

A Thesis Submitted for the Degree of PhD at the University of Warwick

Permanent WRAP URL:

<http://wrap.warwick.ac.uk/149020>

Copyright and reuse:

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it.

Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk



Can routinely collected data be used to inform randomised controlled trial outcomes in oncology?

by
Harriet Paige Mintz

**Thesis submitted for the degree of Doctor of Philosophy
(Health Sciences)**

The University of Warwick, Warwick Medical School

September 2019

Table of Contents

<i>Table of figures</i>	<i>viii</i>
<i>Table of tables</i>	<i>xiii</i>
<i>Dedication</i>	<i>xxii</i>
<i>Acknowledgements</i>	<i>xxiii</i>
<i>Declaration</i>	<i>xxv</i>
<i>Funding</i>	<i>xxviii</i>
<i>Abbreviations</i>	<i>xxix</i>
<i>Abstract</i>	<i>xxxiii</i>
1 CHAPTER ONE: Introduction	1
1.1 Routinely collected data in healthcare	1
1.2 Types of routine clinical healthcare data	3
1.2.1 Administrative data	4
1.2.2 Registry data	7
1.2.3 Administrative data vs. registry data	8
1.3 Healthcare data challenges	8
1.3.1 Regulatory issues	8
1.3.2 Practical issues	9
1.4 The clinical need	11
1.5 Clinical trials and routine data	11
1.5.1 Prostate cancer and the STAMPEDE trial	11
1.5.2 Bladder cancer and the BladderPath trial	14
1.5.3 Clinical trial outcomes	18
1.5.4 Outcome definitions used in this thesis	18
1.6 Rationale for this thesis	20
1.7 Potential positive implications of using routine data for trials	21
1.8 Aims and objectives	23

1.8.1	Aims.....	23
1.8.2	Objectives.....	23
1.9	Thesis outline.....	24
1.10	Thesis summary	24
2	<i>CHAPTER TWO: Shared methodology, methods and further considerations</i>	25
2.1	Shared methodology.....	25
2.2	Shared methods.....	25
2.3	Shared data sources.....	26
2.4	Shared ethics and other approvals	29
2.4.1	Routine data: Hospital Episode Statistics (HES) and the local Hospital Interactions Data (HID).....	29
2.4.2	Routine data – other	30
2.4.3	Clinical trial data.....	31
2.4.4	Hospital clinical noting data	31
2.5	Shared terminology	32
2.5.1	Outcomes and endpoints.....	32
2.5.2	Inferred	32
2.6	Shared bias considerations	32
3	<i>CHAPTER THREE: Direct RCT data collection (retrospective model)</i>	35
3.1	Abstract.....	35
3.1.1	Introduction:	35
3.1.2	Methods:.....	35
3.1.3	Results:.....	35
3.1.4	Conclusion:.....	36
3.2	Declaration	36
3.3	Introduction.....	37
3.3.1	Clinical trial routine data use	37
3.3.2	Neutropenic sepsis.....	38
3.4	Objective	44

3.5	Systematic literature review: Introduction	46
3.5.1	Rationale	46
3.5.2	Aims and objectives	46
3.6	Systematic literature review: Materials and methods	47
3.6.1	Eligibility criteria	47
3.6.2	Information sources and search strategy	47
3.6.3	Study selection	48
3.6.4	Data collection process	49
3.6.5	Data items	49
3.6.6	Risk of bias	49
3.6.7	Summary measures	50
3.6.8	Synthesis of results/additional analyses	50
3.7	Systematic literature review: Results	51
3.7.1	Study selection	51
3.7.2	Summary of the literature identified	52
3.7.3	HSPC vs. CRPC neutropenic toxicity	53
3.8	Systematic literature review: Discussion	58
3.9	Systematic literature review: Conclusion	61
3.10	Use of routine data to evaluate severe neutropenic events in patients administered docetaxel for HSPC and CRPC: Rationale	62
3.11	Materials and methods	63
3.11.1	Approvals	63
3.11.2	Cohorts	64
3.11.3	Data sources and events of interest	68
3.11.4	Data censor	72
3.11.5	Routine data linkage	73
3.11.6	Data preparation	73
3.11.7	Analytical methods	75
3.12	Results	84
3.12.1	Routine data linkage statistics	84
3.12.2	Patient characteristics	84
3.12.3	Cohort one	85
3.12.4	Cohort two	93

3.12.5	Cohort three	96
3.12.6	Cohort four	98
3.12.7	Overall results	101
3.13	Discussion	104
3.13.2	Hypothesis generation	106
3.13.3	Routine data - chemotherapy	107
3.13.4	Routine data - neutropenic events	110
3.13.5	Routine data linkage	113
3.13.6	Algorithm rules	114
3.13.7	HSPC vs. CRPC routine data outcomes	115
3.13.8	Further considerations	116
3.14	Conclusion	117
4	<i>CHAPTER FOUR: Indirect RCT data collection (retrospective model)</i>	<i>119</i>
4.1	Disclosure	119
4.2	Abstract	119
4.2.1	Background	119
4.2.2	Methods	119
4.2.3	Results	120
4.2.4	Conclusion	120
4.3	Introduction	121
4.4	Systematic literature review: Abstract	123
4.4.1	Background and objectives	123
4.4.2	Search methods	123
4.4.3	Results	123
4.4.4	Authors conclusions and implications	123
4.5	Systematic literature review: Aims and objectives	124
4.6	Systematic literature review: Materials and methods	125
4.6.1	Criteria for selecting studies for this review	125
4.6.2	Search methods for identification of studies	125
4.6.3	Exclusion and inclusion criteria	127
4.6.4	Data collection process	128
4.6.5	Reporting	129
4.6.6	Assessment of heterogeneity	129

4.6.7	Risk of bias.....	130
4.7	Systematic literature review: Results and discussion.....	130
4.7.1	Study selection - PRISMA	130
4.7.2	Summary of the literature identified	131
4.7.3	Content analysis – weakness and bias	140
4.7.4	Methodology analysis summary	156
4.8	Systematic literature review: Conclusion	164
4.9	Materials and methods	165
4.9.1	Approvals	165
4.9.2	Participants	165
4.9.3	Data extraction, outcome measures and processing.....	166
4.9.4	Data censoring	168
4.9.5	Requirements to develop a surrogate endpoint.....	168
4.9.6	Analytical methods.....	169
4.10	Results	176
4.10.1	Participants	176
4.10.2	Model development (pilot/feasibility) (cohort 1)	176
4.10.3	Algorithm training (cohort 2)	183
4.10.4	Algorithm validation	187
4.11	Discussion	198
4.12	Conclusion	209
5	<i>CHAPTER FIVE: Direct RCT data collection (prospective follow-up model)</i>	210
5.1	Disclosure	210
5.2	Abstract	210
5.2.1	Background	210
5.2.2	Methods	210
5.2.3	Results	211
5.2.4	Conclusion	211
5.3	Introduction.....	212
5.3.1	Direct clinical trial analysis	212
5.3.2	Chapter rationale	213
5.4	Background.....	214

5.4.1	The dawn of routine data in UK RCTs.....	214
5.4.2	Validating UK RCT routine data	214
5.4.3	Routine data supplementing follow-up in UK RCTs	215
5.4.4	Routine data as the basis for follow-up in UK RCTs	216
5.5	Materials and methods	217
5.5.1	Summary methods	217
5.5.2	Joint methods.....	218
5.5.3	Time to correct treatment (TTCT) analysis.....	218
5.5.4	Data quality assessment	220
5.6	Results	225
5.6.1	Participants	225
5.6.2	Time to correct treatment (TTCT) analysis.....	226
5.6.3	Data quality assessment	227
5.6.4	The framework.....	232
5.7	Discussion.....	234
5.7.1	Time to correct treatment (TTCT) study.....	234
5.7.2	Data quality validation study	236
5.7.3	Implications to RCT conduct.....	238
5.7.4	Further considerations for using routine data in trials	242
5.8	Conclusion	245
6	<i>CHAPTER SIX: Overall discussion and conclusion.....</i>	246
6.1	Discussion.....	246
6.1.1	Routine data concerns	247
6.1.2	Routine data trial potential	257
6.2	Future work	260
6.3	Conclusion	261
7	<i>References</i>	262
8	<i>Appendices</i>	284
8.1	Chapter three	284
8.1.1	PRISMA guidelines (neutropenic events review)	284
8.1.2	Note review confirmed inferred events.....	286
8.1.3	PHE application	286

8.1.4	PHE data variables.....	290
8.2	Chapter four	293
8.2.1	PRISMA guidelines (systematic literature review)	293
8.2.2	Content analyses	294
8.2.3	Interval analyses.....	298
8.2.4	Threshold analyses.....	299
8.2.5	Outcome analyses	300
8.2.6	SRE analysis	301
8.2.7	Outcomes	302
8.2.8	SRE scatter plot	303
8.3	Chapter five	303
8.3.1	Time to correct treatment (TTCT) event coding.....	303
8.3.2	Data quality event coding	305

Table of figures

Figure 1: A simplified data flow for the three main types of routine healthcare data.....	4
Figure 2: A simplified flow of data (dashed lines) from the patient event occurring at the individual healthcare provider level, through to the creation of the multi-provider HES data.....	7
Figure 3: The STAMPEDE MAMS trial, currently with treatment arms A-L.	13
Figure 4: The current standard pathway and NHS guideline targets for the management of NMIBC and MIBC.	16
Figure 5: Example simplified clinical trial patient pathways. A) prostate cancer pathway in the STAMPEDE trial; B) potential disease state transitions from treatment failure with HSPC to CRPC. The red dashed line in A corresponds to the further details in B (I have created this figure based upon the figure previously published by the STAMPEDE trial team (88); C) bladder cancer pathway in the BladderPath trial.	18
Figure 6: The accumulation of longitudinal healthcare data in parallel to the movement through the trial, from randomisation until death or last known follow up.	21
Figure 7: A PRISMA flow diagram showing the stages undertaken to identify the seven papers analysed in the qualitative narrative review.....	51
Figure 8: Across study reported rates of febrile neutropenia.	56
Figure 9: The earliest and latest interaction events identified in the HES and SACT datasets.	63
Figure 10: Venn diagrams illustrating the overlap between study cohorts.....	65
Figure 11: Patients included in the large-scale analyses, cohorts three and four.	66
Figure 12: Cohorts analysed from the STAMPEDE trial.	67

Figure 13: Figure to show the three separate sources of data and their intersections....	68
Figure 14: The potential classification of events from the administration of chemotherapy.	76
Figure 15: The various processes undertaken in the stage one to four cohort analyses.	83
Figure 16: Number of CRPC chemotherapy regimens identified by the HES-only algorithm, prior to inferring missing regimens.	88
Figure 17: The number of neutropenic events identified by the algorithms, prior to inferring missing events.	88
Figure 18: The sensitivity, specificity, PPV and NPV for detecting neutropenic events in the routine data compared to the STAMPEDE trial.	100
Figure 19: The accuracy of detecting STAMPEDE documented events in the HES data (2006 – 2013).	101
Figure 20: Forest plot showing the number of events (n) by the number of patients (N) in each analysis. ORs are displayed, an OR < 1 favours the HSPC group.	102
Figure 21: A PRISMA flow diagram showing the stages of the review.	131
Figure 22: The number of papers published between 2002 and 2019 (until May).	138
Figure 23: The percentage of papers fulfilling each of the 17 criteria analysed during the content analyses.	154
Figure 24: The percentage of the 17 criteria fulfilled by each author.	155
Figure 25: Part 1: Summary of the concepts identified during the full-text review of 12/27 papers analysed in the review; Sx = surgery; Dx = diagnosis; Tx = treatment. RT =	

radiotherapy; HTx = hormone therapy; M1 = metastatic; TNM = tumour, node, metastases. The references correspond to the order of those in table 45.160

Figure 26: Part 2: Summary of the main concepts identified during the full-text review of 9/27 papers analysed in the review; Sx = surgery; Dx = diagnosis; Tx = treatment. RT = radiotherapy; HTx = hormone therapy; M1 = metastatic; TNM = tumour, node, metastases. The references correspond to the order of those in table 45.161

Figure 27: Summary of the main concepts identified during full-text review of the 6/27 papers analysed in the review.162

Figure 28: Diagram illustrating the final cohort selection (cohort four), from the total in the docetaxel comparison in the STAMPEDE trial.166

Figure 29: An example of the 8-weekly clustering of the note review sourced prostate cancer-related events, for one patient, from randomisation (t=0) until death.....179

Figure 30: Clustering of the routine HES prostate cancer-related events (for the patient in figure 29) to identify outcomes of interest (progression = red dots).180

Figure 31: The number of progression outcomes identified by the data source at,181

Figure 32: A Venn diagram showing the number of SREs identified in the trial data, the note review (clinical data) and the routine data.182

Figure 33: Kaplan-Meier analyses for STAMPEDE chemotherapy (C, E) and non-chemotherapy (A, B, D, F) treatment arms, comparing the algorithm output with the STAMPEDE trial MFS.186

Figure 34: Kaplan-Meier analyses for STAMPEDE chemotherapy (C,E) and non-chemotherapy (A,B,D,F) treatments, comparing the routine data outputs with the (A) exact trial PFS and (B) with the trial MFS grouped into the corresponding time-intervals used in the algorithm.....189

Figure 35: Kaplan-Meier analyses (without censor lines) comparing docetaxel (+chemo) with the standard-of-care arm (-chemo). The HES AFS and SRE endpoint are compared to, A) the STAMPEDE MFS, B) the STAMPEDE SREs. The steps in the HES curves = the clustering intervals.....	192
Figure 36: A scatter plot showing the correlation between the HES endpoint and the STAMPEDE, A) MFS and B) FFS.	194
Figure 37: Bland-Altman plots between the HES AFS and the STAMPEDE A) MFS and B) PFS.	196
Figure 38: A summary forest plot. Hazard ratios (HR) (calculated using Cox-proportional hazards models), number of events (n), total number of patients (N) and confidence intervals for each comparison for chemotherapy (CE) and non-chemotherapy (A ± BDF) arms for the different cohorts are displayed.....	197
Figure 39: The study profile for the cohorts utilised in the TTCT analyses.....	219
Figure 40: The PPV for identifying the MIBC patients.	227
Figure 41: The time to definitive treatment for MIBC patients.....	227
Figure 42: The number of events identified in the last five full data years (2013-2017).	229
Figure 43: Accuracy of the routine data to detect the BladderPath primary outcome measure.	232
Figure 44: The methodology developed for the BladderPath trial.....	233
Figure 45: The PRISMA guidelines used to undertake the systematic review.....	285
Figure 46: Project summary and data specification excerpt from the PHE application that I completed for the STAMPEDE routine data.	290

Figure 47: Copy of the document of the variables provided by PHE, upon application. 293

Figure 48: The PRISMA guidelines utilised in the systematic literature review294

Figure 49: A scatter plot of the correlation between the HES detected SREs and the
STAMPEDE detected SREs.....303

Table of tables

Table 1: Routine data concerns for use in trials.	10
Table 2: STAMPEDE trial endpoints.	20
Table 3: Potential positive implications for using routine data for trial conduct.	21
Table 4: Summary of the research stages undertaken during the thesis.	28
Table 5: ICD coding that should be utilised post 01/04/18 to document neutropenic sepsis events.	40
Table 6: The incidence of neutropenic AEs across trials and audits.	44
Table 7: The Medline search strategy that was used to identify articles for the narrative literature review.	48
Table 8: Rationale for the risk of bias investigation.	50
Table 9: Summary of the search results.	53
Table 10: HSPC and CRPC neutropenic event (neutropenia, febrile neutropenia, neutropenic sepsis or infection with neutropenia) incidence values identified in the review studies (terminology as reported within each individual study).....	54
Table 11: Estimated confidence intervals calculated from the audit AE rates, compared to published RCT incidence values (*).	56
Table 12: The earliest and latest interaction events identified in the HES and SACT datasets, provided by PHE.	63
Table 13: The four cohorts used in the analyses.	64
Table 14: Cross-tabulation of STAMPEDE participants and membership in each cohort.	65

Table 15: The data elements extracted from the clinical noting, STAMPEDE data and the routine data.	69
Table 16: Codes identified in the routine data by the algorithms to identify neutropenic admissions.	70
Table 17: OPCS codes identified by the HES routine data algorithms to identify chemotherapy regimens.	71
Table 18: Censor dates used for the analyses.	73
Table 19: Fields utilised from the HES data table for analysis.	74
Table 20: Fields utilised from the SACT data tables for analysis.	74
Table 21: The nine main algorithm rules created due to the lack of clinical details within the routine data.	77
Table 22: Table showing which algorithms utilised which rules.	78
Table 23: The algorithms developed for the different analyses.	80
Table 24: A summary of the analysis completed in each cohort.	82
Table 25: Patient characteristics by cohort.	85
Table 26: The number of patients on HSPC and CRPC chemotherapy regimens and the incidence of neutropenic events.	87
Table 27: The number of events and patients identified using the routine data algorithm compared to the reference note review and the impact of inferring missing chemotherapy regimens. Percentages are rounded to one decimal place. Grey: highlighting the reference values.	89
Table 28: Cohort one analysis HSPC and CRPC rates and the corresponding ORs.	90

Table 29: The incidence of neutropenic admissions from HSPC or CRPC chemotherapy, identified using the HES algorithms (\pm inferral) with the neutropenic event inclusion of ≤ 4 or ≤ 12 weeks after the last chemotherapy event.....	91
Table 30: Detection of neutropenic events.	92
Table 31: The incidence of neutropenic events by the timing of chemotherapy (HSPC, CRPC) identified using the routine data algorithms (varying time intervals for event association, clinical noting validation of missing HES regimens and the SACT analysis), during cohort two analysis.	93
Table 32: Odds ratios comparing the incidence of experiencing a neutropenic event by chemotherapy setting.....	94
Table 33: Assessment of the six extra events that were identified during the inferral analysis.	95
Table 34: ORs comparing the incidence of experiencing a neutropenic event by chemotherapy setting.....	97
Table 35: The ability of the SACT data to detect HSPC STAMPEDE docetaxel regimens..	97
Table 36: ORs comparing the incidence of experiencing a neutropenic event depending on chemotherapy setting.	99
Table 37: The number of events identified in the routine data that were documented by the STAMPEDE trial (all instances of febrile neutropenia, plus grade 3+ neutropenic events).	101
Table 38: Incidence values for the largest cohort for each algorithm.....	102
Table 39: The number of events identified by the algorithm and dataset analysed, including the percentage incidence by timing of chemotherapy administration.....	103

Table 40: The systematic search terms used to conduct this literature review.....	127
Table 41: The inclusion and exclusion criteria for the review.	128
Table 42: Part 1) Summary of findings table for the 27 papers identified.	134
Table 43: Part 2) Summary of findings table for the 27 papers identified.	137
Table 44: The variable coding themes identified from the literature; outcome, procedure, clinical and activity.....	156
Table 45: The methods utilised for algorithm development by the author.	157
Table 46: The number of patients (total cohort size and number of unique/novel patients at the time of the analyses) in each cohort by the stage of analyses.....	166
Table 47: The routine data (HES) fields analysed during model development, to identify prostate cancer-related events.	167
Table 48: A summary of the methods used during the analyses.....	169
Table 49: The events included in the three accuracy calculations, for: missing, inferred and excluded events.	170
Table 50: The events that were identified during the pilot analyses and were included and excluded from the clinical noting data to build the HES-outcome detecting model.	171
Table 51: The rules developed to detect the routine data-derived outcomes.....	173
Table 52: The prostate cancer-related interactions of interest, for the algorithm to detect in the routine data, to flag an outcome of interest.....	173

Table 53: Patient characteristics, split by the analyses cohort (cohort 1 model development N=6; cohort 2 algorithm training N=44; cohort 3 algorithm validation N=46 & 93; cohort 4 algorithm validation N=1,695).	176
Table 54: The accuracy of the HES routine data coding for prostate cancer-related events, since randomisation, for the three different accuracy calculations (missing events, inferred events and excluded events) for four patients.	177
Table 55: The number of progression outcomes identified in the clinical noting that were detected using the clustering of events model. The patient numbers do not necessarily correspond to those in table 54.....	177
Table 56: Analyses into the optimum threshold to identify outcomes of interest. The threshold was chosen as a compromise between sensitivity and PPV. Specificity values were not possible to calculate, nor the NPV, due to an infinite number of false negatives.	180
Table 57: Accuracy of the routine data and the trial data for identifying SREs, in comparison to the note review reference, split by date randomised. The number of SREs correspond to figure 32.	182
Table 58: Reasons for the algorithm-missed/delayed outcomes.	184
Table 59: The number of outcomes identified during the note review (note review PFS), those documented in the trial (STAMPEDE PFS, MFS, FFS, SRE) and those identified using the HES algorithm (HES AFS, SRE), from randomisation, until the mutual censor date.	185
Table 60: The hazard ratios and 95% CIs comparing the HES model output for the HES AFS and SREs, compared to the trial-derived endpoints (FFS, PFS, MFS and SRE) and the note review derived progression (note review PFS).....	185
Table 61: The median (weeks) and 2-year and 5-year event free rate (%) comparing chemotherapy and non-chemotherapy arms using the STAMPEDE endpoints and the HES-derived endpoint.	187

Table 62: The number of outcomes documented in the trial (STAMPEDE PFS, MFS, FFS, SRE) and identified using the HES algorithm (HES AFS, SRE), from randomisation, until censor date.	188
Table 63: The hazard ratios and 95% CIs comparing the HES model output for the HES-derived outcome and HES detected SREs, with the trial-derived endpoints (FFS, PFS, MFS and SREs).	188
Table 64: The median (weeks) and 2-year and 5-year survival (%) comparing chemotherapy and non-chemotherapy arms using the STAMPEDE endpoints and the HES-derived endpoint.	190
Table 65: The number of endpoint events documented in the trial (PFS, MFS, FFS, SRE) and identified using the HES algorithm (AFS, SRE), from randomisation, until censor date.	191
Table 66: The hazard ratios and 95% CIs comparing the HES model output for the routine data derived-outcomes and HES identified SREs, compared to the trial-derived endpoints (STAMPEDE FFS, PFS, MFS and SREs).....	191
Table 67: The median (weeks) and 2-year and 5-year survival (%) comparing chemotherapy and non-chemotherapy arms using the STAMPEDE endpoints and the HES-derived endpoint.	193
Table 68: The correlation coefficients calculated to compare the HES endpoints to the standard STAMPEDE trial endpoints.....	193
Table 69: The algorithm process to extract suspected MIBC patients to enable time to definitive treatment calculation.	220
Table 70: The data sources utilised for the validation study.....	222
Table 71: The number of patients, number of regimens and number of individual events (administrations, radiotherapy fractions), validated by event type.	222

Table 72: Patient characteristics for the TTCT assessment and the data quality analyses for the cohorts.	225
Table 73: Validating the routine data algorithm coding to identify the correct events in a small cohort.	226
Table 74: The total number of events that occurred (identified in the reference), compared to the total number of events that the routine data were able to identify. .	228
Table 75: The number of events identified in the routine data and compared to the reference (2011-2017). The sensitivity is displayed as a percentage (%).	228
Table 76: The sensitivity and PPV estimates of the routine data (2008-2018) coding when events were compared to the reference.	231
Table 77: An example for the BladderPath trial of how additional datasets can aid event identification.....	240
Table 78: The 22 patients that experienced a neutropenic event identified using the HES algorithm with inferral. Whether or not the algorithm could identify these events without inferral is also present. The reason for the missing data is also described, including whether the chemotherapy data could be identified using the SACT data. ...	286
Table 79: Part 1) Content analyses, assessing weakness and bias. * Registry = considered accurate to identify patients but not to identify outcomes; ** If the algorithm was developed a priori (not trained using a sample) then the study was considered to validate on an unseen dataset.....	296
Table 80: Part 2) Content analyses, assessing weakness and bias.* Registry = considered accurate to identify patients but not to identify outcomes; ** If the algorithm was developed a priori (not trained using a sample) then the study was considered to validate on an unseen dataset.....	297

Table 81: Assessment of this thesis study in comparison to the developed weakness and bias criteria from the systematic review.	297
Table 82: Interval analysis conducted to find the optimum time interval to identify outcomes from the prostate cancer-related event clustering.....	298
Table 83: The rationale behind the 8-weekly event clustering that was chosen for the algorithm.	299
Table 84: Further detail into the number of events detected and missed at the five and six event thresholds.....	299
Table 85: Comparative analyses for the five and six event thresholds for the routine data model, compared to the note review (also grouped into the corresponding 8-week interval) and the STAMPEDE trial.	300
Table 86: The difference in the model detected-events (shown in table 85) and the note review detected events, using the 5 and 6 event thresholds and the 8-week intervals.	300
Table 87: Table calculating the accuracy of the routine data to identify SREs. RT = radiotherapy; SCC = spinal cord compression; Blue = note review, HES and trial data; yellow = note review and HES; green = note review and trial data; pink = note review only.	301
Table 88: The number of cohort two outcomes detected, split by treatment, for each endpoint.	302
Table 89: The number of cohort three outcomes detected, split by treatment, for each endpoint.	302
Table 90: The number of cohort four outcomes detected, split by treatment, for each endpoint.	302

Table 91: The codes used to identify the events to calculate time to correct treatment (TTCT) from the routine data.	305
--	-----

Table 92: Codes identified by the algorithm to detect events for the data quality validation.	307
---	-----

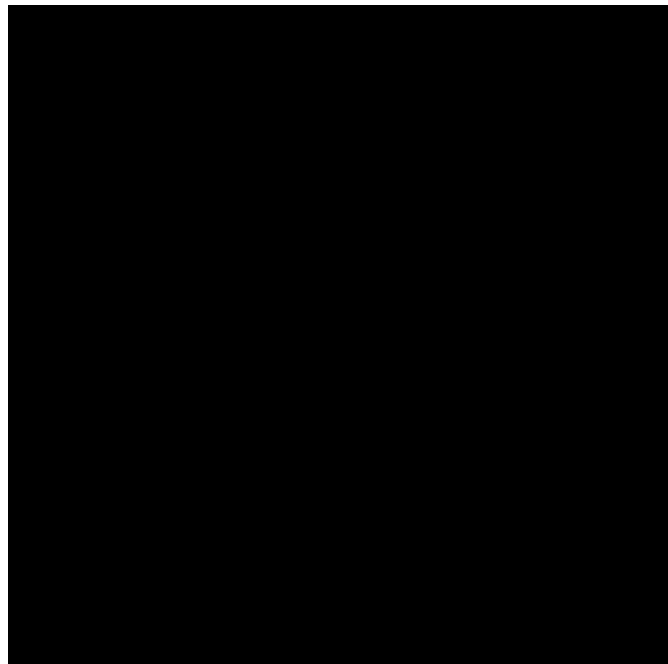
Dedication

I dedicate this thesis to the most determined person I know, my Grandma (Ga). During the completion of this PhD, Grandma has experienced ill health; including, not one, but four fractures. I write this whilst sat next to her in a full leg plaster. Thank you, Ga, for continually listening, supporting me and reading my work, despite it all.

You are a very special lady x

Addendum post submission: Sadly, on the 2nd of January 2020 we lost Grandma, just shortly before my viva. I have always been immensely proud of my Grandma who was the first lady bank manager in the United Kingdom. I think everyone can learn a little something from Ga, be it with her determination, endless strength, generosity, style or her unbroken cheeky smile. She has always supported me unconditionally; there is now an enormous void in our lives. We had such a lot of fun Ga!

You were a huge part of my PhD and now it's done!



Nora Elsie Trixie Howell (Ga) 

Acknowledgements

Firstly, I owe a huge dept of gratitude to my supervisors, Professor Nicholas James and Dr. Helen Parsons. Nick has not only continually supported me with his endless enthusiastic drive but has also had the greatest influence on me; one day I hope that I can reach even a low level of his vast academic and clinical excellence. Without Nick's confidence in me, support, and influence, I would not be taking this incredibly exciting path. Thank you, Nick. I also have extremely special thanks to give to Helen. Helen has not only given me constant unconditional statistical support despite my inexperience but has also provided me with continual mental support, infinite positivity and endless perseverance. Thank you, Helen. I am also particularly indebted to my mentor, Mr Prashant Patel (Pp). Pp has always found time, notably outside of working hours, to guide and support me. I owe much of my future to the influence of Pp, due to his time, advice, total dedication, kindness and clinical exposure that he has given me. I am not only privileged to have made lifelong colleagues in Nick, Helen and Pp, but also friends.

Thank you also to the Medical Research Council Clinical Trials Unit at UCL team, including past and present STAMPEDE trial team members; Matt Sydes, Claire Amos, Mary Rauchenberger, Chris Brawley, Jo Calvert, Carly Au, Fiona Ingleby, Mel Gannon and Max Parmar.

Thank you also to the Birmingham team (at the Queen Elizabeth hospital and the University); especially to Amandeep Dosanjh. Aman, we made a great informatics team. Thank you for all your work and support. Thank you also to Mehreen Mahmood (if you want something done, ask Mehreen), Ana Hughes, Manushri Jain, Felicity Evison, Alicia Jakeman, Ann Pope and all the team in the Clinical Research Facility office and not to forget, Naj and Shirley in the café for providing me with coffee. Also, thank you to the PHE team, making data applications bearable; Rachael Brannan, Luke Hounsborne, Haiyan Wu and Sean McPhail. I have thoroughly enjoyed being part of this multi-institutional collaborative team, with whom I will continue to co-author and work with. In addition, a very special thank you to all the patients involved, especially to those in the STAMPEDE trial.

I also want to give specific thanks to the University of Warwick, for my last seven years of education, since leaving school and for funding my studentship. Also, thank you Sam (Plumb) and of course, Sean (Barrett) in the Medical school.

Finally, and of the greatest importance to me, where thanks are not enough. To my Mom (Sue), Dad (Bern), Boyfriend (Matt), Brother (Ell) and Grandma (Ga); the five of you have held me together with your unconditional support and encouragement.

Declaration

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. I, Harriet Paige Mintz, declare that it has been composed by myself and has not been submitted in any previous application for any degree.

The work presented (including data generated and data analyses) was carried out by the author, except in the cases outlined below:

List of data provided and/or analysis carried out by collaborators:

Datasets provided by collaborators:

- The chapter three and four data were provided by: Routine healthcare data (National Health Service Digital (NHSD) (previous application), Public Health England (PHE) (application completed by myself)); trial data (the Medical Research Council Clinical Trials Unit at University College London (MRC CTU at UCL)).
- The chapter five data were provided by: Routine healthcare data (the informatics department at University Hospitals Birmingham Queen Elizabeth Hospital (UHB QEH)); clinical reference data (a surgical reference compiled at the UHB QEH and a radiotherapy reference extracted by the UHB QEH).

Analyses carried out by collaborators:

- Amandeep Dosanjh: Calculated patient characteristic summary statistics for the data quality analyses cohorts, described in the main text (Chapter 5, table 72).
- Christopher Brawley: Calculated patient characteristic summary statistics for the largest multi-site analyses, described in the main text (Chapter 3, table 25).

- Carly Au: Calculated the total STAMPEDE trial patient numbers (chapter 4, figure 28).

Parts of this thesis have been published by the author:

Published conference abstracts (not including others presented, but not published via a citable source but including those accepted but pending presentation and those presented but not yet citeable):

1. National, centralised hospital datasets can inform clinical trial outcomes in prostate cancer: A pilot study in the STAMPEDE trial (1)

Conference: Genitourinary cancers symposium, American Society of Clinical Oncology (ASCO)

Dates: 16 - 18th February 2017

Location: Rosen Shingle Creek, Orlando, Florida, United States of America (USA)

Published: Journal of Clinical Oncology

2. The feasibility of using routinely collected data to inform outcomes in randomised controlled trials (RCT) in oncology

Conference: Trials using cohorts & routine health data: International symposium on their efficiency and analysis

Dates: 15th May 2019

Location: Wellcome Collection, Euston Road, London, United Kingdom (UK)

Published: Trials (pending)

3. Retrospective evaluation of neutropenic admission events in metastatic or high-risk hormone-sensitive prostate cancer (HSPC) patients having docetaxel chemotherapy upfront or for castrate-resistant prostate cancer (CRPC) in STAMPEDE

Conference: European Society for Medical Oncology (ESMO) congress

Dates: 27th September - 1st October 2019 (to be presented on the 30th September)

Location: Fira Gran Via, Barcelona, Italy

Published: Annals of Oncology (pending)

4. **Routinely-collected hospital datasets can be used to identify endpoints predictive of overall survival outcomes in randomised controlled trials (RCT): a prostate cancer study within the STAMPEDE protocol (NCT00268476)**

Conference: International Clinical Trials Methodology Conference (ICTMC)

Dates: 6th - 9th October 2019 (to be presented on the 9th October)

Location: Hilton Brighton Metropole, Brighton, UK

Published: Trials (pending)

Published/submitted papers:

1. **Development and validation of a follow-up methodology for a randomised controlled trial, utilising routine clinical data as an alternative to traditional designs: a pilot study to assess the feasibility of use for the BladderPath trial**

Journal: Pilot and feasibility studies

Status: Submitted, pending peer-review

Funding

A studentship to undertake this PhD was awarded to Harriet Paige Mintz by the University of Warwick, Warwick Medical School.

Abbreviations

A&E	Accident and emergency
ADT	Androgen Deprivation Therapy
AE	Adverse event
AFS	Activity-free survival
ASCO	American Society of Clinical Oncology
BCG	Bacillus Calmette-Guérin
CanCORS	Cancer Care Outcomes Research and Surveillance study
CaP	Cluster Randomized Trial of PSA Testing for Prostate Cancer
CASP	Critical Appraisals Skills Programme
CHAARTED	Chemohormonal Therapy Versus Androgen Ablation Randomised Trial for Extensive Disease in Prostate Cancer
CI	Confidence interval
CND	Clinical noting data
COSD	Cancer Outcomes and Services Dataset
CPRD	Clinical Practice Research Datalink
CRF	Case report form
CRN-VDW	Cancer Research Network Virtual Data Warehouse
CRPC	Castrate-resistant prostate cancer
CT	Computerised tomography
CTCAE	Common Terminology Criteria for Adverse Events
DBS	Disclosure and Barring Service
DID	Diagnostic Imaging Dataset
DNA	Deoxyribonucleic acid
DOB	Date of birth
DPIA	Data Protection Impact Assessment
DPR	Danish Pathology Registry
Dx	Diagnosis
EHR	Electronic healthcare record
ESMO	European Society for Medical Oncology
FFS	Failure-free survival
G-CSF	Granulocyte colony-stimulating factor

GCP	Good clinical practice
GDPR	General Data Protection Regulation
GP	General practice
GPRD	General Practice Research Database
HDR UK	Health Data Research UK
HES	Hospital Episode Statistics
HID	Hospital Interactions Data
HR	Hazard ratio
HSCIC	Health and Social Care Information Centre
HSPC	Hormone-sensitive prostate cancer
HTx	Hormone therapy
ICD	International Statistical Classification of Diseases
ICECaP	International Intermediate Clinical Endpoints in Cancer of the Prostate
ICTMC	International Clinical Trials Methodology Conference
ID	Identification
IQR	Interquartile range
ISD	Information Service Division
JCO	Journal of Clinical Oncology
KAT	Knee Arthroplasty Trial
KM	Kaplan-Meier
LABKA	Clinical Laboratory Information System
M	Metastases
M0	Non-metastatic
M1	Metastatic
MAMS	Multi-Arm-Multi-Stage
MDT	Multiple disciplinary meeting
MFS	Metastases-free survival
MHRA	Medicines and Healthcare Products Regulatory Agency
MIBC	Muscle invasive bladder cancer
MRC CTU at UCL	Medical Research Council Clinical Trials Unit at University College London
MRI	Magnetic resonance imaging
N	Node

N/n	Number
N+	Node positive
N0	Node negative
NA	Not applicable
NCR	National Cancer Registry
NCRAS	National Cancer Registration and Analysis Service
NE	Neutropenic event
NHS	National Health Service
NHSD	National Health Service Digital
NICE	National Institute for Clinical Excellence
NIHR	National Institute for Health Research
NIHR HTA	National Institute for Health Research, Health Technology Assessment
NMIBC	Non-muscle invasive bladder cancer
NOC	Not otherwise classifiable
NPV	Negative predictive value
NR	Not reported
NS	Not specified
NX	Nodal status cannot be measured
ONS	Office of National Statistics
OPCS	Office of Population Censuses and Surveys Classification of Interventions and Procedures
OR	Odds ratio
OS	Overall survival
PAS	Patient Administration System
PbR	Payment by Results
PDS	Personal Demographics Service
PET	Positron emission tomography
PFS	Progression-free survival
PHE	Public Health England
PPV	Positive predictive value
PRISMA	Preferred reporting items for systematic reviews and meta-analyses
PROM	Patient reported outcome measure
PSA	Prostate specific antigen

RCT	Randomised controlled trial
RD	Routine data
REACT	Randomised Evaluations of Accepted Choices in Treatment
REC	Research Ethics Committee
RTDS	National Radiotherapy Dataset
SACT	Systemic Anti-Cancer Therapy Dataset
SAE	Serious adverse event
SAIL	Secure Anonymised Information Linkage
SD	Standard deviation
Se	Sensitivity
SEER	Surveillance, Epidemiology, and End Results
SOC	Standard of care
SOP	Standard operating procedures
Sp	Specificity
SRE	Skeletal-related event
STAMPEDE	Systemic Therapy in Advancing or Metastatic Prostate Cancer: Evaluation of Drug Efficacy
SUS	Secondary Uses Service
Sx	Surgery
T	Tumour
TD	Trial data
TTCT	Time to correct treatment
TURBT	Transurethral resection of bladder tumour
TX	Tumour cannot be measured
Tx	Treatment
UHB QEH	University Hospitals Birmingham Queen Elizabeth Hospital
UK	United Kingdom
USA	United States of America
VA	Veterans Administration
WoS	Web of Science
WTCRF	Wellcome Trust Clinical Research Facility
Y/N	Yes/No

Abstract

Introduction: Randomised controlled trials (RCT) have supplemented standard data collection with routine healthcare data. However, no RCTs in the United Kingdom have been conducted solely using routine data in oncology or secondary care. This thesis was undertaken to assess methods to enable the replacement or supplementation of standard RCT data. I present examples of routine data follow-up in two clinical settings: prostate and bladder cancer.

Methods: Routine healthcare datasets were validated against reference patient data (for example, trial data and clinical noting), for their ability to identify trial outcomes of interest. Models were developed to algorithmically identify these outcomes from the routine data. Outcomes included: toxicity (serious adverse events), disease progression, treatments and the last known follow-up interaction.

Results: Models were developed enabling the identification of outcomes of interest from the routine data, for example sepsis admissions and trial non-survival endpoints, for example, progression. This enabled the estimation of uncollected trial case report form (CRF) events, which subsequently have become of interest. I developed a novel routine data-derived endpoint, which correlated with standard trial endpoints, enabling estimation of treatment effects from routine data. I also developed a method to validate the feasibility of using routine data as the basis for oncology trial follow-up.

Discussion: The nature of the routine data meant that models had to be developed to enable identification of some events of interest *indirectly*. Although routine data quality was shown to be improving, techniques had to be implemented, for example, through data querying, to ensure integrity, accuracy and relevance. Routine data can provide a robust method of trial data collection but needs to be used in combination with other data sources, such as, standard trial data or clinical noting.

Conclusion: I propose that routine data are a feasible source of trial outcomes; however, each individual outcome requires validation.

1 CHAPTER ONE: Introduction

1.1 Routinely collected data in healthcare

There remains a clear therapeutic need for improved interventions in the field of oncology (2-4). Clinical trials, however, can be complex, resource intensive, expensive to perform and funding from the government and other sources is limited (5). It has been proposed that randomised controlled trials (RCT) are experiencing '*increasingly prohibitive costs of conducting adequately powered studies*' (6). There are also other limitations, including lack of research into rare diseases (7) and often an inability to retrospectively answer new research questions due to inadequate case report form (CRF) outcomes reported for the new hypotheses. Therefore, clinical progress is restricted, and the rate of therapeutic advances has reduced. An article published in a series from the journal *Nature* (2014, (8)) investigated the success rates for interventional drugs in trials; they identified that drug productivity may be lower than previous estimates and concluded that, '*adaptive clinical trial designs and improved methodologies and greater flexibility with alternative surrogate endpoints are areas in which this productivity can be improved*' (8).

There are many methods to enhance trial design, for example, the use of a Multi-Arm-Multi-Stage (MAMS) platform, whereby a simultaneous assessment of many different interventions can be undertaken against a single control arm (9). In addition, pre-existing dataset analysis is another potential way to provide an enhanced framework to establish results (9). Routinely collected healthcare data may therefore, offer an efficient alternative for trials (10).

Back in 2000 it was recognised that despite RCTs being widely accepted as the optimum method to assess outcomes, sometimes they have limited use. For example, if the RCT is not feasible (11). It was thus proposed in 2000 that routinely collected data could be used to alleviate these problems. However, it was also stated that there was '*insufficient information on patients' conditions [in the routine data] at discharge to enable a comparison of outcomes*' (11). Eleven years on from 2000 (2011), an article, appropriately titled, '*Extracting value from chaos*', explained that our digital footprints

are increasing (12) and alongside this increase in data, it has been proposed that healthcare data may now '*resemble extremely closely the unknown distribution of the clinical phenomenon of interest*' (13). Over the last few decades there has been an increase in the size and complexity of data and its value in healthcare is increasingly being comprehended and accepted. It has been stated that, '*the secret of human disease may lurk under the vast ocean of big data, waiting [for] us to decode and understand them*' (14).

However, as initially identified in 2000, insufficient outcomes remain a limitation of such resources and therefore, analyses often require *data mining* techniques, or *indirect* identification through the use of alternative coding. Insufficiently reported routine data outcomes include, treatment and disease events such as non-survival endpoints, for example, disease progression (see further details in 1.5.3) and patient reported outcomes, such as quality of life data. One definition of data mining is the use of methods to analyse large volumes of data (15), but often these models do not get adapted for clinical use. In addition, coding errors and missing data are limitations (15). There are further challenges in harvesting these data (12), both practical and regulatory. Further practical issues include the difficulty of extracting useful information from a large, un-cleaned, exponentially growing, raw dataset with a potential lag between data collection and processing (16) and regulatory issues include, safeguarding, privacy, ownership and governance. These issues are discussed further below in section 1.3.

Despite these limitations, the overriding opinion is that there is vast potential for these data to lead to improved patient outcomes. For example, leading to a longer and improved quality of life, through enhancing research and development with tools and algorithms to improve trial design (16). This potential can be seen with the timely set-up of the Health Data Research UK (HDR UK) initiative. HDR UK was established in April 2018 (17) and one of its aims is to develop *better, faster and more efficient clinical trials* (18) using health data, to aid recruitment and follow-up.

Hence, this thesis aimed to extract trial outcomes from routine healthcare data, by establishing methods to create and evaluate models, whilst ensuring feasibility for clinical use, with respect to both practical and regulatory issues.

1.2 Types of routine clinical healthcare data

Although the terminology is often used interchangeably, the following are the main types of routine healthcare data; 1) electronic healthcare records (EHR), 2) administrative data (collected for a primary function other than research) or claims data (for example, health insurance claims, particularly common in the United States of America (USA)) and 3) registry data (figure 1) (19).

EHRs (20) are the electronic version of the patient clinical noting, which often are used to create the other datasets (both the administrative and the registry data). EHRs contain unstructured and structured data. Structured data are derived from defined fields and coding schemes, whereas unstructured data can be derived from sources such as free-text clinical noting which may be scanned (21). Free-text may be searchable, for example, a word document. However, scanned noting may be non-searchable and therefore it is difficult to extract data (22). Clinical noting data are now largely electronic but unstructured and often electronically unsearchable.

Administrative or claims data are collected primarily for non-research uses, for example, for reimbursement service purposes. In contrast, registry data are collected to evaluate specific populations, usually upon diagnosis of a disease. Clinical registry data can also be derived from other sources, including manual clinician input into a website-based interface, patient reporting systems, or these data can be derived from medical devices and services (23). For example, the National Radiotherapy Dataset (RTDS), which derives data from the radiotherapy treatment machine, through an oncology management system (figure 1) (24).

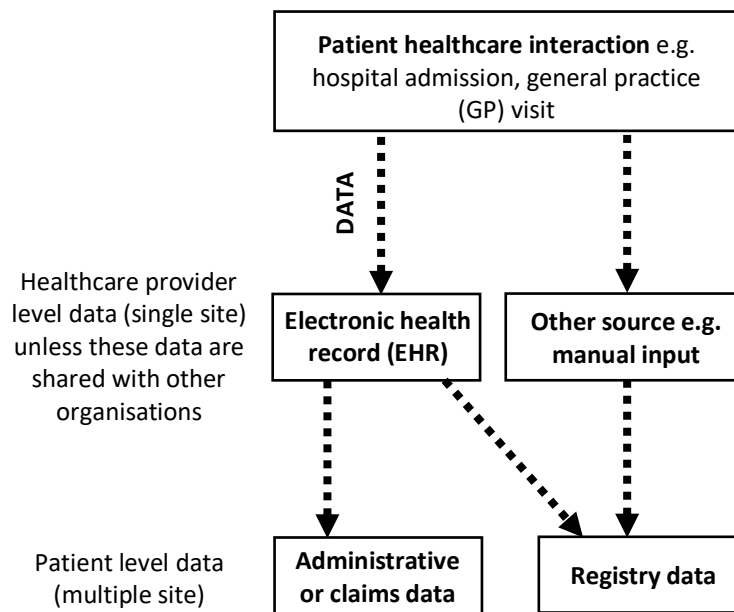


Figure 1: A simplified data flow for the three main types of routine healthcare data.
Dashed line = flow of data.

The nomenclature for these data sources is complex, due to the varying data derivation sources and interchangeable definitions used within the literature. Thus, this thesis will refer to routinely collected healthcare data as, 1) administrative and 2) registry data (figure 1).

1.2.1 Administrative data

There are many different administrative datasets in the United Kingdom (UK). Examples include, datasets collected in both primary care (general practice (GP)-derived) and secondary care (hospital-derived). A primary care example is, the Clinical Practice Research Datalink (CPRD) (25, 26) which collects primary care patient interactions throughout the UK. The secondary care data are managed by National Health Service (NHS) Digital (NHSD) in England (the Hospital Episode Statistics (HES)) (27), the Secure Anonymised Information Linkage (SAIL) Databank in Wales (enabling linkage of multiple primary and secondary care datasets) (28, 29) and the Information Service Division (ISD) in Scotland (30, 31). In the UK, all non-private care occurs through the NHS and due to this, NHS secondary healthcare interactions are documented by NHSD, SAIL or ISD data.

Due to this research being undertaken in a secondary care setting in England, the administrative database investigated here was the Hospital Episode Statistics (HES). In addition, a local dataset was used during the non-trial analyses. These local data are the hospital data that are returned to form the HES data (figure 2), so can be assumed to be a HES equivalent. In addition, a registry dataset was utilised, which is discussed below (section 1.2.2).

1.2.1.1 The Hospital Episode Statistics (HES)

HES data are collected for every NHS inpatient, outpatient and accident and emergency (A&E) visit in England. The database contains disease and procedure fields populated by International Statistical Classification of Diseases (ICD) (32) and Office of Population Censuses and Surveys (OPCS) Classification of Interventions and Procedures (33) codes respectively.

The HES contain clinical, patient, administrative and geographical information but contain no data on disease or treatment outcomes. For example, the date that a patient's disease has worsened (disease progression) or that the treatment is no longer working (treatment failure) is not documented. Taking the example of progression, whilst subsequent treatments for a progression event (for example, chemotherapy) should be reported in the HES, this progression outcome would not be reported. This means that often clinical trial outcomes cannot be *directly* isolated from these data.

Although not documented in the HES data, time to cancer progression is often a key trial endpoint, which functions as a surrogate for overall survival (OS) (time to death) (34). OS is often the preferred indicator of clinical benefit. It is possible to collect death data using NHSD collected sources and many RCTs currently use such data sources (35). However, other endpoints (non-survival) are reached prior to OS (34) and are therefore desirable. Delaying the time to these endpoints being reached is often of clinical benefit to patients and therefore these events are documented in trials (34) (details on cancer outcomes can be seen in section 1.5.3 and 1.5.4).

The HES dataset is vast; data collection began in 1987 with the collection of admitted patient care records after a national programme was rolled out (36). Outpatient data

were collected from 2003 and A&E data from 2007. In addition, methods of collecting these data have changed dramatically since 1987; initially annual data collection was undertaken, but currently HES are released monthly and undergoes a two week cleaning process, meaning more processed data can be accessed more frequently (27).

As shown (figure 1), data from the EHR are used to form the administrative data, hence clinical HES data are derived from EHR coding (figure 2). Data from the EHR are formulated into an administrative system (the hospital Patient Administration System (PAS)), through translation of events into the diagnosis and procedure codes. Translation of the unstructured clinical event data into the PAS is undertaken manually by clinical coding staff at a single healthcare provider level, in contrast to structured data which can be automatically incorporated into datasets. This manual translation can lead to errors (37) and systematic differences in the coding between hospital trusts (for example, see section 3.3.2.2).

Extracts of these data are sent on a frequent basis to the Secondary Uses Service (SUS), where these data are stored and processed (outside of the NHS trust that these data were derived from) to become the HES data, or where they are prepared for secondary uses. One major use of these data is, Payment by Results (PbR) (38); these data can be prepared to contain financial information so hospitals can be paid for the care they provide (figure 2) (27). These data are also used for audit purposes, for benchmarking and policy making for service improvement. Users include national bodies and regulators, such as the Department of Health and the National Institute for Clinical Excellence (NICE), governmental departments and researchers (39).

Processed HES data are frequently used for research, including, to estimate costs (40, 41), epidemiology (42, 43), to identify risk factors (44, 45) and in validation studies of data quality (46). In addition, these data are less frequently used to identify non-survival events, such as progression (47).

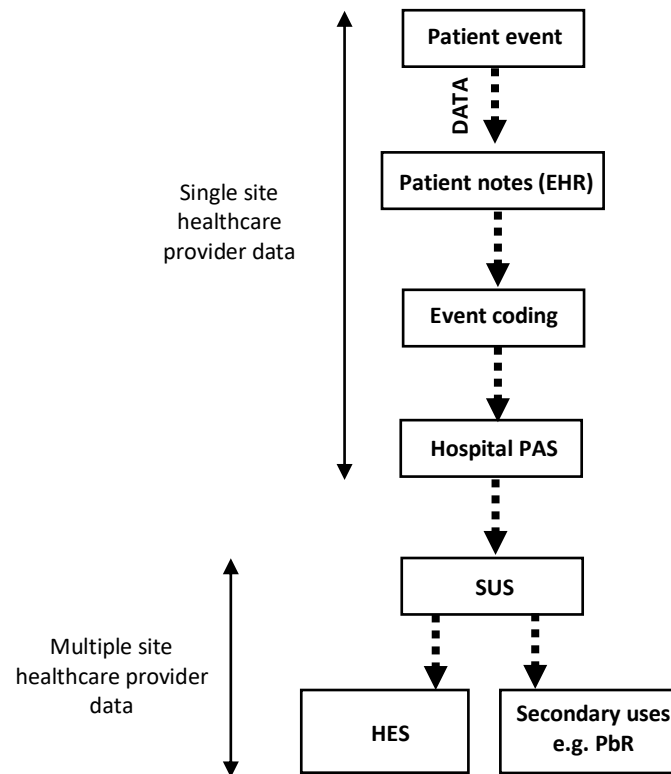


Figure 2: A simplified flow of data (dashed lines) from the patient event occurring at the individual healthcare provider level, through to the creation of the multi-provider HES data.

The process of cleaning the SUS data is vast and involves four main steps, including, editing organisation codes (provider code mapping), identifying clear data errors (data cleaning), deriving extra information such as geographical data from the coding (derivations) and removing duplicates (duplicate removal) (48).

1.2.2 Registry data

There are many different disease registries, however, the National Cancer Registry in the UK collects data for all NHS patients (through an opt-out regime) diagnosed with cancer. It is managed by the National Cancer Registration and Analysis Service (NCRAS) which is part of Public Health England (PHE) (49). The NCRAS is involved in various projects, for example, the Cancer Outcomes and Services Dataset (COSD) (50), the RTDS (24) and the Systemic Anti-Cancer Therapy (SACT) dataset (51, 52). The main registry data used in this thesis were the Systemic Anti-Cancer Therapy dataset (SACT) (51, 52).

1.2.2.1 The Systemic Anti-Cancer Therapy (SACT) dataset

From May 2014, all NHS trusts that deliver anti-cancer regimens were required to document data in the SACT registry. SACT collects therapy data for all adult and paediatric haematological or solid tumours (51). For example, oral chemotherapy, standard chemotherapy, immunotherapy and steroid regimens are required to be collected (53).

1.2.3 Administrative data vs. registry data

The main difference between registry and administrative data is the purpose of collection. Registry data are collected for research/audit purposes, whereas administrative data are often collected for other purposes (for example, payment). This has implications on the quality and completeness of these data. Hospitals can be fined for incorrect reporting of administrative data, as inaccurate measures of performance can be generated and the hospitals may not get paid for the care they provide (37). In contrast, because the primary function of registry data is not for administrative purposes, this can lead to sporadic reporting. An example of this is the method of data collection for recurrence, in the COSD. These data are manually collated, often in multiple disciplinary meetings (MDT), which is labour intensive (54) and results in missed cases (23).

1.3 Healthcare data challenges

There are many practical and regulatory challenges that accompany using routine healthcare data. All of these challenges needed to be considered during the conduct of this thesis (table 1).

1.3.1 Regulatory issues

Regulatory concerns arise when using routine data in trials (table 1). Issues include, privacy, security and consent. In parallel to the evolution of electronic medical record usage alongside the dawn of valuable data repositories, there was a rise in privacy and security concerns (55, 56). Hence, in 1996 it was acknowledged that a universal security policy was lacking in the healthcare industry (57). Therefore, the following year, the Caldicott Report was released, which is still in use today to define clinical governance

rules (58). In addition, on the 25th May 2018 the General Data Protection Regulation (GDPR) was initiated to strengthen data protection (59). It applies to all *personal data*, defined as, data with the ability to identify any individual by identifiers (directly or indirectly). The GDPR documents seven key principles: 1) lawfulness, fairness and transparency, 2) purpose limitation, 3) data minimisation, 4) accuracy, 5) storage limitation, 6) integrity and confidentiality (security) and 7) accountability (60). The studies in this thesis were undertaken during the implementation of the GDPR.

Although all of these regulations are absolutely necessary to protect patient privacy, it has since been proposed that these regulations, that were created to protect us, have now created large, valuable but inaccessible data repositories or ‘silos’ (61). Such regulations are thus proposed to have halted innovation and these regulations previously created to protect us from harm, are actually now leading to harm (61). The Times 2017 stated that, *‘the NHS could transform diagnosis and treatment if only we overcame our distrust and let it share patient data’* (62).

1.3.2 Practical issues

Practical issues with utilising routinely collected data for trials, include, data accuracy (quality), outcome availability (limited outcomes), cost (63), linkage, timeliness of collection and bias (table 1). Routine healthcare data are often seen to have quality issues, where the coding is incorrect or missing (46, 64, 65). Consideration of this is of utmost importance during these studies, as trial data integrity is vital. For example, estimates of treatment effects may be changed upon loss to follow-up, leading to biased and invalid results (66). Despite these routine data inadequacies, such data have been seen to be improving (46).

As discussed above (section 1.2.1.1), clinical variables of interest are often not available in routine healthcare datasets, for example, progression outcomes (10) (see section 1.5.3 and 1.5.4 for further details on cancer outcomes). Thus, it has been stated that for trials that need strictly defined endpoints, standard dedicated follow-up is required for patient safety and quality endpoint reporting (10). Hence, it was proposed that routine healthcare data (registry data) can be used for collecting baseline variables only, and cannot be used for comparative outcomes research (10).

The Academy of Medical Sciences and the Association of the British Pharmaceutical Industry released a report in 2017 expressing the need for new trial endpoints (2). The report stated that new endpoints seek to ‘*accelerate the development of anti-cancer treatments*’ whilst aiming for a strong overall survival association but also reflecting patient priorities (2). It was here proposed that routine clinical data could be used to identify or validate existing endpoints or for long-term tracking of responses. However, it was also highlighted that prior to this, considerable work is required to develop useful and reliable outcomes that are compatible across different datasets (2).

Another potential limitation is data timeliness; data are required that are timely in order to use such a resource for clinical trial analyses (63). Data must be available with a suitable delay from when the event occurred but also available at frequent intervals from the data provider. Consideration of the frequency of data transfers from routine data providers and the delay in event collection in the routine data are both vital for trial design. Costs also have to be considered; it is believed that routine data follow-up techniques will reduce costs (63). However, assessments of the routine data methods compared to the standard trial follow-up techniques are required to confirm this. Linkage rates are also a concern, if trial data cannot be linked to the routine sources, then loss to follow-up will occur and outcomes will be missed. Another limitation is that the routine data studies can also be biased impacting the study integrity (see section 2.6) (table 1).

	Aspect	Notes
Practical	Outcome availability	Can the outcome required by the trial be identified in the routine data?
	Accuracy	Is data quality high enough, compared to standard trial data?
	Linkage	Can these data be linked to the trial identification numbers (ID) or loss to follow-up will occur?
	Cost	Does using a routine data framework reduce the costs compared to standard follow-up?
	Timeliness	Are these data recent enough to perform trial analyses? Are recent events accessible?
	Bias	Does the use of routine data lead to biased studies; for example, ascertainment, design, selection and temporal issues?
Regulatory	Privacy, security and consent	Are all regulatory aspects in place to enable access and use of the routine data?
	Data retention	Can the routine data be kept long enough for an adequate audit trail for the trial?

Table 1: Routine data concerns for use in trials.

Despite these concerns, in addition to the timely HDR UK initiation (section 1.1), the National Institute for Health Research (NIHR) are currently calling out for 'data-enabled trials' (deadline September 2019). Grants are being offered for projects, such as those within this thesis. This highlights the timeliness of this research (67), despite the practical and regulatory issues.

Due to these data practicality and regulatory concerns, this thesis aimed to assess if feasible methods could be developed to 1) replace/supplement existing clinical follow-up within the well-established prostate cancer RCT, STAMPEDE (section 1.5.1) and 2) if feasible, enable routine data to be used as the basis of follow-up for the new bladder cancer RCT, BladderPath (section 1.5.2). The background behind the trials and outcomes of interest are documented below.

1.4 The clinical need

Cardiovascular disease has been the leading cause of death worldwide; however, it was predicted that by 2020 cancer would become the leading cause of death (68). In fact, many countries already record more deaths from cancer (69) and therefore, strategies need to be developed to reduce this incidence. This thesis investigated two different urological malignancies, prostate and bladder cancer, both resulting in high morbidity (the amount of disease in a population) and mortality worldwide (70, 71).

1.5 Clinical trials and routine data

Routine data are increasingly being used in RCTs (chapter 5, section 5.4). The routine data can be linked by unique patient identifiers, such as the NHS number, to the trial ID to enable patient level data to be analysed (63). The STAMPEDE trial (72) and the BladderPath (73) trial are exemplar RCTs used within the studies presented in this thesis.

1.5.1 Prostate cancer and the STAMPEDE trial

Prostate cancer is the most common male cancer in the UK (74) and in 2008 it was suggested to be the '*sixth leading cause of cancer death among men worldwide*' (70). In 2008 there were 899,000 new cases and 258,000 deaths due to prostate cancer, worldwide (70). Due to the ageing of the population and growth, by 2030 these figures are expected to rise to 1.7 million new cases and 499,000 new deaths per year.

However, it was estimated that in 2012 prostate cancer already accounted for 1.1 million cases and 307,000 deaths, so is fast out-growing its projected estimates (70, 75).

Prostate cancer can progress (worsen) from the confines of the prostate gland (non-metastatic (M0) with or without the local lymph nodes of the pelvis) to distant (metastatic, M1) lymph nodes out of the pelvis or to other distant M1 sites (76, 77). These distant M1 sites can include the viscera (organs) and the bone (78). Skeletal related events (SRE) may occur if the cancer has progressed to the bone. An example of an SRE is spinal cord compression, a complication of bone cancer; SREs are discussed in detail in section 1.5.4. The cancer may also be recurrent, where the disease can return, when there was previously a response to treatment (79). This can result in the need for treatment change (79).

The STAMPEDE (72) (Systemic Therapy for Advancing or Metastatic Prostate cancer: Evaluation of Drug Efficacy) trial is the largest interventional prostate cancer trial worldwide and is testing multiple treatments in advancing or metastatic disease, in comparison to a single control treatment (the standard-of-care, SOC). The STAMPEDE trial uses a MAMS design (9, 80, 81) to allow simultaneous assessment of multiple treatments against the SOC (9, 82) (figure 3). The STAMPEDE SOC is currently hormone therapy (Androgen Deprivation Therapy, ADT), with or without radiotherapy (RT) in relevant subgroups, plus docetaxel chemotherapy or abiraterone hormone therapy.

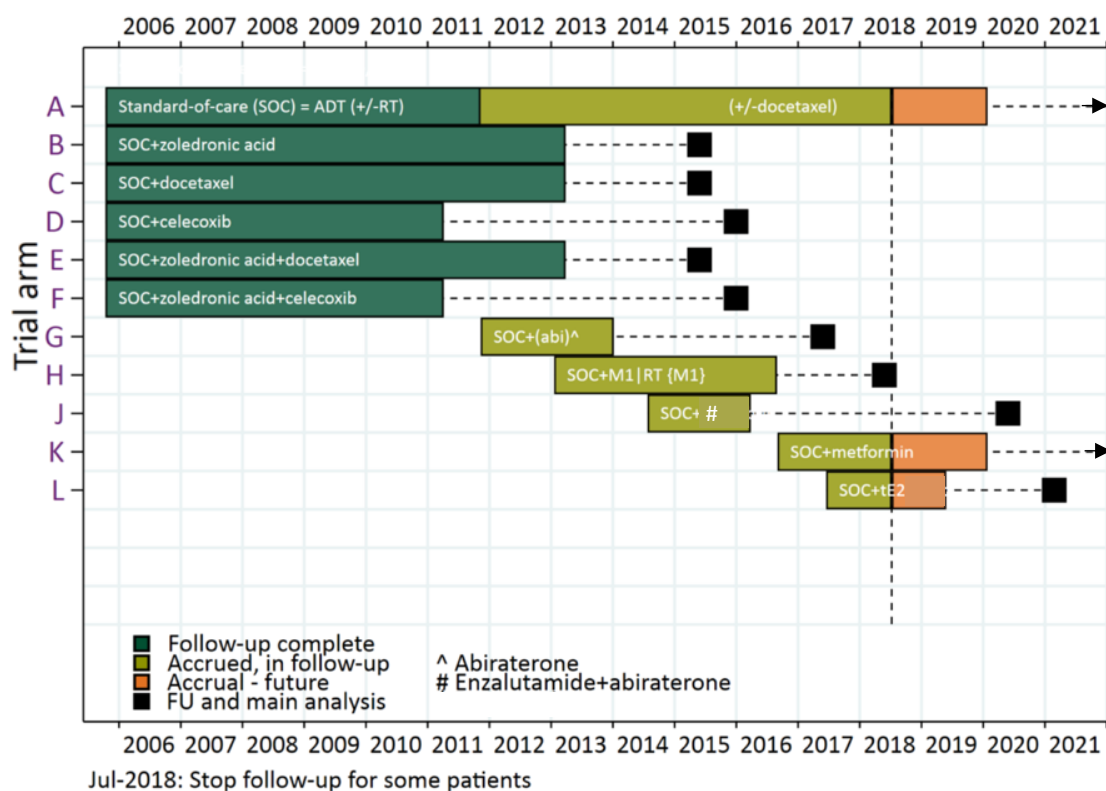


Figure 3: The STAMPEDE MAMS trial, currently with treatment arms A-L.
This figure was extracted and edited from the STAMPEDE protocol (72).

STAMPEDE has recruited over 11,000 patients with high-risk disease localised to the prostate gland, the lymph nodes or metastatic sites, or relapsing disease after initial treatment to the prostate (72). If a patient is diagnosed with this, they may be eligible to enter the STAMPEDE trial (figure 3) (72).

STAMPEDE has so far evaluated or is evaluating: the SOC (arm A) (see description above); SOC plus zoledronic acid bisphosphonate treatment (83) (arm B); docetaxel (83) (arm C); celecoxib (84) (arm D); zoledronic acid plus docetaxel (83) (arm E); zoledronic acid plus celecoxib (84) (arm F); abiraterone (85) (arm G); M1 radiotherapy (86) (arm H); enzalutamide hormone therapy plus abiraterone (arm J); metformin (arm K) and a hormone therapy patch (72) (arm L). The STAMPEDE treatment arms investigated during these studies were arms A-G (figure 3).

A STAMPEDE eligibility criterion (arms A-K, not arm L), includes the intention to treat with long-term hormone therapy (androgen deprivation therapy, ADT). Hence, when patients enter the trial, they have not had prior exposure to long-term hormone therapy and therefore, the patients on these arms are assumed to be recruited with hormone sensitive prostate cancer (HSPC) (72). This means that the prostate cancer should respond to hormone therapy. Patients can be administered the STAMPEDE SOC treatments (arm A, control arm) or be recruited to the other experimental arms which includes various treatments (arm B-L) (figure 3).

The patients can develop recurrent disease after initial treatment. At this point, if the patient is no longer responding to hormone therapy, castrate resistant prostate cancer has developed (CRPC) (87) (see section 3.3.2.3 for further details). In the order of severity (most severe, to least severe), this progression to the CRPC state can include: visceral metastases (internal organs), bone metastases including an SRE, bone metastases without an SRE, CRPC disease without metastases or lymph-node metastases alone (including, the non-metastatic local lymph nodes of the pelvis or distant metastatic lymph nodes) (88). Continual progression can occur and death from prostate cancer or non-prostate cancer causes is possible in any state, regardless of the severity (88). Subsequent treatments may be suitable and the recurrent and non-recurrent groups are continually followed-up by the trial until death. An example of a STAMPEDE trial patient pathway can be seen in section 1.5.2.1.

1.5.2 Bladder cancer and the BladderPath trial

In 2012 it was estimated that bladder cancer was the 9th most common cause of cancer worldwide when combining both sexes (75). The aim of the study presented within this thesis was to determine the feasibility of using routine data for follow-up for the BladderPath trial for bladder cancer (ISRCTN35296862) (73). However, the initial data analyses were conducted in a non-trial patient cohort.

When a patient is newly diagnosed with bladder cancer they may be eligible to enter the BladderPath trial (73). In summary, the rationale for the trial is as follows: the outcomes for patients with bladder cancer have not changed significantly for many decades and one hypothesised reason for this is the delay from diagnosis to the definitive treatment

(the first management intervention) in the more severe disease subgroup, muscle invasive bladder cancer (MIBC) (73). The BladderPath trial (Image Directed Redesign of Bladder Cancer Treatment Pathway) is an RCT assessing a redesigned patient pathway, with the aim to fast track the correct patients to the most appropriate treatment, to reduce delays and improve outcomes (73, 89).

The standard pathway is that patients have a flexible cystoscopy to diagnose bladder cancer, which is followed by rigid cystoscopy under general anaesthetic during which piecemeal excision of tumour can be undertaken - the transurethral resection of bladder tumour (TURBT) (73). Diagnosis can include non-muscle invasive bladder cancer (NMIBC), where the cancer has not invaded the muscle layer, or the more severe form, MIBC, where the cancer has invaded the muscle layer (90). TURBT not only acts as a diagnostic tool, but it is also used for NMIBC treatment, as the tumour can be potentially completely excised during the investigative process. However, if MIBC is present, further treatment is necessary: typically cystectomy, chemotherapy, or radiotherapy or combinations of all three, or palliative care (73). In addition, both NMIBC and MIBC can become recurrent and if so, further treatment is additionally required; NMIBC can also progress to MIBC disease. Multiple recurrences can occur and hence, where appropriate, additional treatments can be given.

In the MIBC subgroup, the TURBT is not only potentially redundant but it may also be damaging. For example, TURBT often incorrectly under-stages MIBC patients as NMIBC (91) leading to delayed MIBC treatment. It is also possible that the invasive procedure within the bladder, with the cystoscope, could lead to dissemination of the tumour (92). Furthermore, imaging post-TURBT can be challenging due to surgical artefacts (for example, damaged tissue) (73) leading to incorrect treatment choices. Service evaluation data show that delays to MIBC treatment are common (93, 94).

The current NHS guidelines for the time from receiving a GP referral to starting treatment for cancer is 62 days, with a second target of 31 days from confirmed diagnosis to starting treatment (95). For both MIBC and NMIBC, TURBT is considered the first treatment, despite the lack of efficacy of TURBT for MIBC. Therefore, although 31 days is the target from diagnosis to treatment, MIBC patients have a further delay to

starting the *appropriate* treatment (figure 4). Varying delays have been identified, including, a study with a further median time from TURBT to cystectomy of 7 weeks (50 days) (96) and another study showing that 90% of patients did not receive definitive treatment within 62 days. The maximum treatment delay was 80 weeks (560 days) from referral (93).

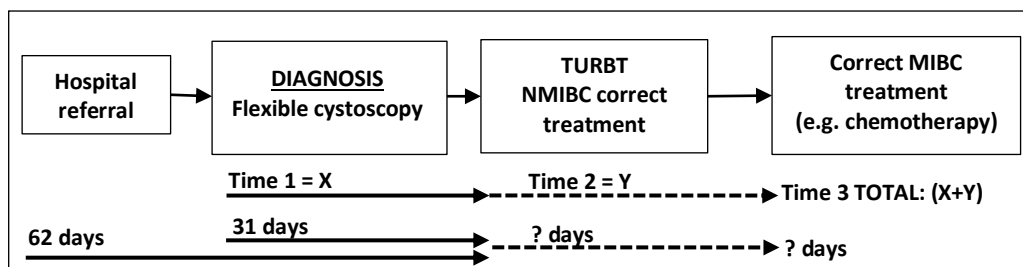


Figure 4: The current standard pathway and NHS guideline targets for the management of NMIBC and MIBC.

Continuous black line = time from hospital referral or diagnosis (where treatment plans are created) to treatment for NMIBC. Dashed line = the additional time to appropriate treatment for MIBC patients. X, Y = the true value for time 1 and 2 respectively.

Thus, BladderPath is testing the hypothesis that NMIBC and MIBC patients could be separated at diagnosis, using Magnetic Resonance Imaging and biopsy at flexible cystoscopy (in particular patient subgroups) to replace the TURBT. Different pathways could then be followed sooner, with the aim to observe better outcomes in the MIBC subset. These patients previously identified as having poorer prognosis, potentially due to delays (96, 97), would be fast-tracked directly to the correct treatment (73). An example of a BladderPath trial patient pathway can be seen in section 1.5.2.1.

1.5.2.1 Prostate and bladder cancer patient pathways

An example patient pathway for a patient with prostate cancer entering the STAMPEDE trial and a patient with bladder cancer entering BladderPath can be seen in figure 5.

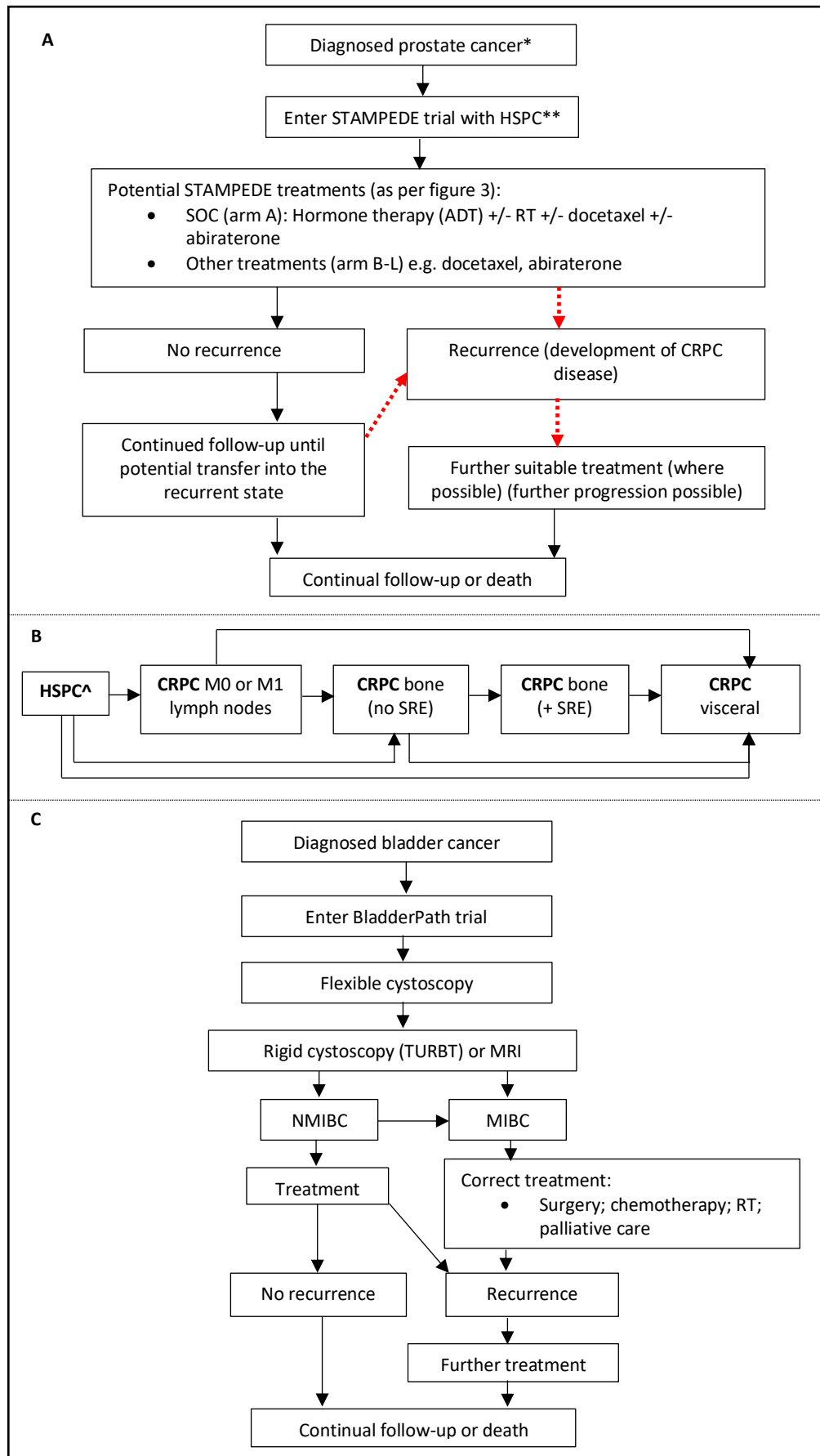


Figure 5: Example simplified clinical trial patient pathways. A) prostate cancer pathway in the STAMPEDE trial; B) potential disease state transitions from treatment failure with HSPC to CRPC. The red dashed line in A corresponds to the further details in B (I have created this figure based upon the figure previously published by the STAMPEDE trial team (88); C) bladder cancer pathway in the BladderPath trial.

* Diagnosed with high-risk locally advanced disease localised to the prostate gland, the lymph nodes or metastatic (M1) sites, or relapsing disease after initial treatment to the prostate; HSPC = hormone sensitive prostate cancer; CRPC = castrate resistant prostate cancer; ** = one eligibility criterion includes the necessity to treat with long-term hormone therapy, hence, are required to have HSPC at randomisation; ADT = androgen deprivation therapy (hormone therapy); RT = radiotherapy; ^ = Patients can be randomised into the STAMPEDE trial with M0 or M1 HSPC. Upon hormone therapy treatment failure in these groups, CRPC is developed; NMIBC = non-muscle invasive bladder cancer; MIBC = muscle invasive bladder cancer.

1.5.3 Clinical trial outcomes

After patients enter oncology trials, outcomes are often collected in paper or electronic case report forms upon trial visits. Outcomes of interest include OS (time from randomisation into the clinical trial, until death of any cause) and other non-survival surrogates/proxies which aim to predict the OS but happen sooner than death. Alternative non-survival endpoints are useful for MAMS trials, as they enable more expedient assessment of trial treatments (see section 1.5.4).

Prostate cancer outcomes include, disease progression, treatment failure, metastases (72) and skeletal-related events (SRE). Despite the use of such endpoints, in prostate cancer trials, no clinical endpoints have been accepted for advanced disease. Metastases-free survival (MFS) was the first (2017) accepted robust surrogate for OS but only in localised prostate cancer (98) (see section 1.5.4). These events are detected objectively (imaging), biochemically (a blood laboratory result change) or symptomatically (a change in symptoms) within the STAMPEDE trial (72).

Other commonly collected outcomes in CRFs include, toxicity (adverse events (AE), serious adverse events (SAE)) and quality of life (72). During event reporting in STAMPEDE, it is only mandatory to collect data for the first event.

1.5.4 Outcome definitions used in this thesis

Failure-free survival (FFS): Within STAMPEDE, FFS is defined as the time to failure, where failure is the first of: death from prostate cancer, biochemical failure, local prostate failure, lymph node failure or distant metastatic spread (99). FFS can be

detected objectively, biochemically or symptomatically as discussed in section 1.5.3. To assess biochemical failure, a unique threshold prostate specific antigen (PSA) blood test marker is calculated for each patient, depending on the lowest value between randomisation and week 24. When the PSA is confirmed to have reached the threshold, biochemical failure (treatment failure) is confirmed as the date of the first increased PSA value. The FFS is then calculated as the time from randomisation to the first event. If no event is present, these data are censored at the last known visit (72) (table 2).

Progression-free survival (PFS): This is determined to be the time to when the disease has worsened. Within STAMPEDE, progression is defined as, death from prostate cancer, local failure, lymph node failure or distant metastases and time to this first event from randomisation is defined as the PFS. If no event is present these data are censored at the last known visit (72, 99) (table 2).

Metastases-free survival (MFS): This is defined as the time to when the disease has spread. Within STAMPEDE, events of interest include, any death (mortality by any cause), distant metastases or progression of metastases. The time to the first event is defined as the MFS. If no event is present these data are censored at the last known visit (72, 99) (table 2).

Overall survival (OS): This is defined as the time to mortality by any cause. Time to this event from randomisation is the OS, this is the primary outcome for most oncology trials. If no event is present these data are censored at the last known visit (72, 99) (table 2). Cause-specific survival includes the cancer-related deaths only.

Skeletal related events (SRE): The bone is usually the first site of metastases in prostate cancer. Two in three patients with cancer will develop bone metastatic disease, hence, skeletal activity can act as a surrogate for disease progression (100, 101). Within STAMPEDE, these events of interest include; pathological fracture, spinal cord compression, requirement for radiotherapy to the bone or requirement for surgery to the bone. If one of these events are identified, progression should be confirmed, and a progression form completed for this. If no event is present these data are censored at the last known visit (72).

Endpoint	Prostate cancer death	Non-prostate cancer death	Biochemical failure	Local failure	Lymph node failure	Distant metastases
FFS	✓	✗	✓	✓	✓	✓
MFS	✓	✓	✗	✗	✗	✓
PFS	✓	✗	✗	✓	✓	✓
OS	✓	✓	✗	✗	✗	✗

Table 2: STAMPEDE trial endpoints.

In addition, SREs are used as an endpoint but are not included in the table.

Time to correct treatment (TTCT): This is the primary outcome for the intermediate stage of the BladderPath trial. The correct treatment depends on the nature of the disease and the TTCT is the time from randomisation to the correct treatment (73).

Adverse events (AE) and serious adverse events (SAE): AEs occur where a treatment has an adverse effect, for example exacerbation of an illness, an increase in an event, a condition detected after trial treatment initiation, or a symptom that worsens following the trial drug (72). Further to this, SAEs are defined as any adverse event that is serious, for example, with death, life-threatening consequences, hospitalisation and persistent disability (72).

1.6 Rationale for this thesis

Routine data have frequently been used within clinical trial frameworks (102). However, in 2015 Professor Nicholas James identified that there was a lack of work into validating and where possible, making clinically useable methods to utilise these data within the oncology trial setting. This thesis was conceived to develop and validate novel methods to use routine data, using Professor James' clinical trials as exemplar data. The long-term ambition was to supplement or replace follow-up in an existing oncology RCT and to develop a novel clinical trial framework for a new RCT using routine data solely for follow-up.

The following three studies interlink and hence were completed largely in parallel to validate different data sets and data items within routinely collected sources. This aimed to determine the feasibility of using such techniques.

This rationale is based upon the idea that, when enrolling into a trial (such as STAMPEDE and BladderPath), patients can consent for long-term information to be accessed from routine sources (72). Therefore, these rich routine data can be accessed, by linkage to the trial ID (figure 6). However, in order to extract meaningful outcomes data (as discussed above) for trials, models have to be created.

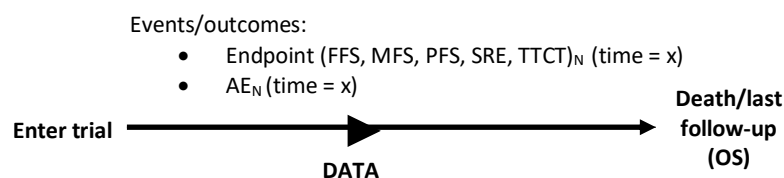


Figure 6: The accumulation of longitudinal healthcare data in parallel to the movement through the trial, from randomisation until death or last known follow up.

N = number of events as multiple events can occur; x = an unknown time, unique to each patient and event. FFS, MFS, PFS, SREs, TTCT and AEs are defined in section 1.5.4.

1.7 Potential positive implications of using routine data for trials

If routine data can be used to identify trial outcomes, either to supplement standard data collection techniques or to replace standard methods, alongside the concerns proposed in table 1, there are many potential positive implications (table 3). The main overall implication linked to all of the aspects shown in table 3 is patient benefit.

Aspect	Implications
New hypotheses	Can hypotheses be explored that previously were not possible using standard trial data?
Validation	Can standard CRF events be validated using routine data?
More events	Can additional events be identified using routine data, that were lost to follow-up using standard trial data? Or extra events that could flag at risk patients?
Reduced burden	Can routine data reduce burden on site staff reporting events and patients visiting for follow-up?
Reduce costs	Can routine data reduce the costs compared to running a trial with standard data collection techniques?
Increase timeliness	Could timeliness of event collection be increased using routine data?

Table 3: Potential positive implications for using routine data for trial conduct.

Hypothesised positive implications include:

- Events could be investigated that were not previously known to be significant. Previously these events would not have been collected in the pre-defined follow-up CRFs and it would not be possible to retrospectively acquire these data for large cohorts (to be investigated in chapter 3) (table 3).
- Traditionally collected trial CRF events could be validated (to be investigated in chapter 3 and chapter 4) (table 3).
- Acquiring trial outcomes from routine resources may reduce burden on site staff reporting events and on patients visiting for frequent follow-up assessments (103, 104). A reduced number of follow-up visits over a trial with long follow-up may be desirable to patients and enhance recruitment (table 3).
- The use of routinely collected data may reduce the cost (103, 105) of conducting the trial due to a proposed reduction in resources that are required (table 3). This is dependent on what additional resource is required to utilise the routine data (103).
- Loss to follow-up may be reduced. Events usually lost to follow-up may be identifiable in the routine data, for example, events previously missed due to recall bias or those occurring at another hospital (to be investigated in chapter 3, chapter 4 and chapter 5) (table 3).
- The timeliness of event collecting may increase. The feasibility of near real-time data collection may enhance event capture, compared to standard data collection at interim trial visits. Especially as these visits can occur at progressively more distant intervals (explored in chapter 5) (table 3).

Ultimately the aim is to enhance clinical trials. Hence, more interventions would potentially become available to patients, with the overall aim to directly improve patient quality and length of life (OS).

1.8 Aims and objectives

1.8.1 Aims

- To develop clinically useable instruments, that utilise routinely collected healthcare data, to detect outcomes for use in trials. The aim is to: 1) assess the use of routine data as the basis of follow-up for a new RCT and if this is identified as feasible, set up a framework to do this, 2) identify if existing follow-up can be replaced or supplemented for an existing RCT and create models to identify events.
- To answer the thesis question, *'Can routinely collected data be used to inform randomised controlled trial outcomes in oncology'?*

1.8.2 Objectives

Direct methodology (retrospective model) (chapter three)

- Determine if it is possible to perform analyses for events that were not routinely collected in the trial CRFs.
- Establish the quality of the routine data compared to the standard trial data and clinical noting data.
- Illustrate the potential data use by investigating a clinical hypothesis using the model.

Indirect methodology (retrospective model) (chapter four)

- Identify in the literature if trial non-survival endpoints have previously been developed using routine data.
- Determine if it is possible to identify trial outcomes not explicitly collected by routine data.
- Establish data quality compared to traditional trial data and clinical noting data.
- Develop and validate an algorithm to identify events of interest for a trial.
- Reproduce major trial analyses using the model.

Direct methodology (prospective model) (chapter five)

- Identify how routine data have been used to conduct oncology trials previously.
- Validate the quality of the routine data.
- Develop an algorithm to identify events of interest.
- Develop a framework for the BladderPath trial data collection.
- Acquire routine data for the trial.

1.9 Thesis outline

Hence, in line with the aims (section 1.8.1) and objectives (section 1.8.2), this thesis presents three main chapters, 1) retrospectively assessing if analysis can be undertaken using routine data to answer questions not previously possible using standard data collection and if possible conduct such analyses; 2) retrospectively assessing if it is possible to *indirectly* identify outcomes not collected in the routine data and if feasible try to perform survival analyses to replicate previous trial analyses; 3) investigating if routinely collected data can be used to identify events that are *directly* identifiable in these data. The aim was to develop a framework to utilise routine data as the basis of data collection as this has not been done in a secondary care or oncology RCTs previously.

These studies were completed to create an overarching knowledge set to determine feasibility and where possible, implement the use of routine data to, 1) replace/supplement STAMPEDE trial follow-up, 2) enable conduct of the BladderPath trial using routine data as the basis of follow-up.

1.10 Thesis summary

Reviews of the literature confirmed that no studies had been undertaken to identify oncology trial surrogate outcomes to perform trial time to event analyses, to identify treatment effectiveness. Including, no urology trials using routine data to identify trial non-survival endpoints. I also believe that no RCTs have been designed to perform follow-up using routine data as the basis of follow-up, in oncology or a secondary care setting. To accomplish this, there were shared approaches between the studies which are outlined in the next chapter (chapter 2).

2 CHAPTER TWO: Shared methodology, methods and further considerations

As discussed in section 1.9 the three studies interlink but the shared approaches across the three studies are presented below. The methods are described as the tools used to undertake the research and the methodology as the broader strategy and rationale that impacts the methods. Hence, the methods are part of the overall methodology. In addition, shared data sources, approvals, terminology and bias considerations are shown.

2.1 Shared methodology

This thesis undertook both primary and secondary research. The primary research includes the studies where I collected and performed the research myself. The secondary research includes the literature reviews, where the primary analyses were previously conducted by another researcher (106). All three analyses chapters (3-5) include both primary and secondary research (table 4).

Both qualitative and quantitative approaches were used and hence, a mixed methods study was adopted. Firstly, qualitative approaches were taken, which were then followed by the quantitative approaches. The qualitative approach refers to where I sought to, *'come to terms with the meaning, not the frequency'* of data (107). For example, by analysing patient case histories, or published literature. In contrast, the quantitative approach refers to where data were analysed statistically.

2.2 Shared methods

All studies involved the validation of the routine data against a reference and algorithm development to identify events. Where note review was undertaken, data were extracted in Microsoft Excel designed data collection tools. Where systematic literature reviews were undertaken, databases were screened and Endnote Web (108) was utilised to manage the references. No second reviewers were utilised for these reviews due to a lack of resource. However, advice was given by H. Parsons and the University of Warwick librarian. Algorithms were developed to identify outcomes in the routine data, and these outcomes were compared to reference data sets to perform sensitivity analyses. In all studies the algorithms were developed using the software R (109) and RStudio (110).

The algorithmic design and analyses were checked by Dr Helen Parsons. Specific details of individual research methods are detailed in the appropriate chapter (chapter 3-5).

2.3 Shared data sources

Three data sources were analysed throughout the research. The reference ('gold standard') was considered to be the hospital clinical noting (where available) and the clinical trial data in the absence of clinical noting (table 4):

1. **Routinely collected healthcare data:** Administrative data: Hospital Episode Statistics (HES) or local Hospital interactions Data (HID) and registry data: Systemic Anti-Cancer Therapy dataset (SACT).
2. **Clinical trial data:** The STAMPEDE trial case report form (CRF) reported data (reference).
3. **Hospital clinical noting:** Patient clinical noting and radiotherapy machine generated data (reference).

Although the three data sources were shared, not all datasets were used for each study. For example, the SACT data were only used within chapter 3.

2.3.1.1 Temporal data considerations

During the conduct of these analyses, there were temporal changes in the datasets. Temporal changes affected all three sources of data. Routinely collected data undergoes changes, for example in the way events are coded and accuracy. To ensure these changes were reflected in the work, these data were validated against the reference sources and analysed by year.

The clinical trial data were also affected by temporal changes. The case report form (CRF) designs can change over time and different variables can be collected. Hence, the relevant CRF data needed to be extracted for each analysis. The clinical noting data also were also affected by temporal changes in data collection; the transition from paper to electronic clinical noting. The single site University Hospitals Birmingham, Queen Elizabeth Hospital (UHB QEH) where the clinical noting data originated, became one of

the first outpatient departments in the United Kingdom to record clinical data electronically. Hence, this enabled access to the majority of patients clinical noting via the computer system. However, I undertook training to enable me to request paper noting for the early patients analysed.

To ensure that the three data sets were comparable, where appropriate, I used data freezes to enable comparison. This is because these three data sets are documented at different times through the patient's disease trajectory. That is, after the patient event has occurred, the clinical noting is the first available, then the trial data or the routine data.

Stage	Data source	Events of interest	Type of analyses	Publications
Direct methodology (retrospective design) (chapter 3)				
Model development and validation	<ul style="list-style-type: none"> Reference 1: Clinical trial data Reference 2: Clinical noting Routine data: HES, SACT 	Neutropenic admissions, chemotherapy, hormone-state	<ul style="list-style-type: none"> Case series Retropective cohort study Secondary data analysis 	<ul style="list-style-type: none"> Presented at <i>Trials using cohorts and routine health: international symposium on their efficiency and analysis</i> (conference poster) – to be published in <i>Trials</i> To be presented at the European Society for Medical Oncology (ESMO) annual meeting (poster) – conference abstract to be published in <i>Annals of Oncology</i> Full journal manuscript in preparation
Model use	<ul style="list-style-type: none"> Routine data: HES, SACT 			
Indirect methodology (retrospective design) (chapter 4)				
Model development				
Model training	<ul style="list-style-type: none"> Reference 1: Clinical trial data 			<ul style="list-style-type: none"> Genitourinary symposium, American Society of Clinical Oncology (ASCO) (poster presentation) – published in the <i>Journal of Clinical Oncology (JCO)</i>
Model validation (single-site, small-scale)	<ul style="list-style-type: none"> Reference 2: Clinical noting Routine data: HES 	Prostate cancer related-events (randomisation to death), including trial endpoints)	<ul style="list-style-type: none"> Case series Retropective cohort study Secondary data analysis 	<ul style="list-style-type: none"> Presented at <i>Trials using cohorts and routine health data: international symposium on their efficiency and analysis</i> (conference poster) – to be published in <i>Trials</i>
Model validation (multi-site, large-scale)	<ul style="list-style-type: none"> Clinical trial data STAMPEDE Routine data: HES 			<ul style="list-style-type: none"> To be presented at the International Clinical Trials Methodology Conference (ICTMC): <i>Routinely-collected hospital datasets can be used to identify endpoints predictive of overall survival outcomes in randomised controlled trials (RCT): a prostate cancer study within the STAMPEDE protocol</i> (NCT00268476) (oral) – conference abstract to be published in <i>Trials</i> Full journal manuscript in preparation (to be submitted)
• Direct methodology (prospective design) (chapter 5)				
Model development and validation	<ul style="list-style-type: none"> Reference 1: clinical noting Reference 2: radiotherapy data Routine data: HID 	Individual treatment events, censor event (last known event)	<ul style="list-style-type: none"> Retropective cohort study Secondary data analysis 	<ul style="list-style-type: none"> Submitted pending peer review to <i>Pilot and Feasibility studies</i> Presented at <i>Trials using cohorts and routine health: international symposium on their efficiency and analysis</i> (conference poster) – to be published in <i>Trials</i>

Table 4: Summary of the research stages undertaken during the thesis.

2.4 Shared ethics and other approvals

Due to these studies using patient level, personally identifiable data, approvals to access and analyse these data were required for all three data sources (routine data, trial data and clinical noting). The approvals required can be seen below.

2.4.1 Routine data: Hospital Episode Statistics (HES) and the local Hospital Interactions Data (HID)

The HES data were analysed from two different sources via two different applications. Initial feasibility studies were undertaken using HES data previously acquired for the STAMPEDE trial from the Health and Social Care Information Centre (HSCIC) (becoming NHS Digital, NHSD in 2016 (111)). After feasibility was determined, analyses were undertaken using multi-site data acquired from Public Health England (PHE). I completed the application for these data (appendix section 8.1.3 and 8.1.4), which took sixteen months from the initial contact with PHE (27/06/17) to receiving these data (08/11/18). This specific sub-study also obtained Research Ethics Committee (REC) approvals (NHS Health Research Authority West Midlands Edgbaston Research Ethics Committee, REC reference: 04/MRE07/35) and is consequently an ancillary study within the STAMPEDE protocol (section 17.4 of the STAMPEDE protocol) (72).

When entering the STAMPEDE trial, the patients consent to follow-up using routine data, enabling us to access their routine data. These routine data were sent from PHE to the data controller, the Medical Research Council, Clinical Trials Unit at University College London (MRC CTU at UCL). Hence, I acquired an honorary contract with UCL to enable me to access these data within the Data Safe Haven (DSH). I have also signed the STAMPEDE delegation log to enable me to perform analyses for the trial. Conditions of the contracts include, updated Good Clinical Practice (GCP), information governance training, the completion of the standard operating procedures (SOP) including tests after each module, and General Data Protection Regulation (GDPR) training, which was required during the completion of these studies; hence, all of the above were completed and passed, where examined.

The local Hospital Interactions Data (HID) for the BladderPath feasibility work were acquired from the University Hospitals Birmingham, Queen Elizabeth Hospital (UHB QEH) informatics department. The work was registered as an audit on the UHB systems and these data were transferred internally within the hospital from the informatics department. I also completed a Data Protection Impact Assessment (DPIA) which was sent to the information governance team at UHB QEH. The BladderPath protocol (73) also documents the data feasibility study (section 10 and 17 of the protocol). Both the STAMPEDE trial and the BladderPath trial also have the relevant information regarding data access available in the trial patient information sheets.

Initially, the HSCIC (now NHSD) sent monthly HES extracts to the UHB QEH informatics department for linkage to the STAMPEDE trial ID. The STAMPEDE-HES linked datasets were analysed at UHB QEH after this linkage process. In contrast, these data from PHE were sent to the MRC CTU at UCL in one retrospective drop, already linked to the STAMPEDE trial ID, so no further linkage was required. The MRC CTU at UCL securely provided the required linkers (STAMPEDE ID, NHS number, date of birth (DOB) and the date the patient entered the trial), to enable data to be sent for events six-months prior to randomisation to the last available routine data interaction. In order to be linked by PHE, a PHE linkage requirement included that the patient was registered with cancer in the registry records and had a cancer record in the HES data. Manual checks on valid NHS numbers were undertaken by PHE prior to extraction of the individual eligible patient records from the cancer registration data. NHS number rules were used to extract patients with valid NHS numbers. The patients with eligible events and a correct NHS number were linked to the STAMPEDE trial ID, the direct identifiers were stripped, and these data returned securely to the MRC CTU at UCL for analyses.

2.4.2 Routine data – other

The PHE data application which I completed enabled the access to HES, National Radiotherapy Dataset (RTDS), Systemic Anti-Cancer Therapy dataset (SACT) and cancer registration data. These SACT data were used in addition to the HES during chapter 3. These SACT data were also linked to the STAMPEDE trial ID by PHE.

2.4.3 Clinical trial data

STAMPEDE is run according to GCP and the Declaration of Helsinki, with relevant regulatory and ethical approvals (9, 81, 83, 84) (STAMPEDE registration numbers: MRC PR08, ISRCTN78818544, NCT00268476). As mentioned in section 2.4.1, this specific sub-study also obtained Research Ethics Committee approvals and is consequently an ancillary study within the STAMPEDE protocol.

As with the requirements to access the linked HES data, I signed the delegation log for the trial to access STAMPEDE data. Paper trial case report forms (CRF) were utilised initially and these were extracted and re-filed by site data managers to ensure safety (of both archived and unarchived CRFs). For larger analyses, I completed two trial data release requests to the MRC CTU at UCL, to apply for the electronic STAMPEDE CRF data to be sent to the UHB QEH. This was necessary due to the volume of patient data required from multiple sites. This was sent to UHB QEH via a secure transfer, to ensure integrity and confidentiality of these data. All data transfers to the UHB QEH were signed off by the head of research and development governance. Further to this, the GCP exams, and mandatory yearly UHB training enabled me to access these data; I passed all of the modules.

2.4.4 Hospital clinical noting data

I obtained a UHB QEH honorary contract to access data at the hospital. Again, an updated GCP certificate, frequent information governance training and other mandatory training sessions were required. In addition, a disclosure and barring service (DBS) check was undertaken. Audits were also registered on the clinical audit system, to access the hospital clinical noting data for the STAMPEDE patients and the non-trial bladder cohort. The Wellcome Trust Clinical Research Facility (WTCRF) at UHB QEH initially oversaw the use of these data and booked particular computers to use. The honorary contract also enabled visits to outpatient clinics and inpatient procedures, all providing an understanding of the clinical background and data acquisition for the studies undertaken.

2.5 Shared terminology

Terminology was shared across the chapters and further details can be seen below in section 2.5.1 and 2.5.2.

2.5.1 Outcomes and endpoints

Various clinical trial outcomes/events were being investigated throughout this project, for example, adverse events (AE) and non-survival time to event endpoints. The terminology therefore is defined within this thesis as: all endpoints are outcomes but not all outcomes are endpoints. Hence, within this thesis an endpoint is defined as an outcome/event.

2.5.2 Inferred

The terminology *inferred* was used within this thesis where non-statistical assumptions were made to account for missing routine data. Instead, clinical justifications were used. For example, in one analysis, for patients who had severe neutropenic events but no chemotherapy details in the routine data, it was assumed that these events were chemotherapy induced. This is because clinical experience leads us to believe, in this population, the majority of severe neutropenic events are triggered by chemotherapy.

2.6 Shared bias considerations

Different types of bias are present within these three studies. This includes selection, sampling, ascertainment, temporal, design and detection bias (112).

Sampling bias may arise due to the small single-site studies undertaken within chapter 3, 4 and 5. Clinicians at different trusts will recruit different patients and in addition, the coding may be different across sites. Hence, single-site analyses may not be representative of the population. There are also issues due to the reporting limitations of the routine data used (HES, SACT); The HES and the SACT are only reported in England and NHS patients. Hence, patients receiving private care and those outside of England are excluded from the analyses.

Due to data not being collected for patients receiving private care, this may influence the cohorts, such as by socioeconomic status. Within chapter 3 (neutropenic event study), if a patient is admitted for a severe neutropenic event, this should be to an NHS emergency department. Hence, the exclusion of private cases should have minimal impact. These patients may be receiving different patterns of care and excluding these patients may bias the results.

Selection bias is also present within these studies. For example, patients within clinical trials may be receiving a better quality of care and hence, this may impact the coding which may not be representative of the whole population. If the patients were in the STAMPEDE trial, these have been selected as are eligible for the trial. Hence, patients not meeting the STAMPEDE inclusion criteria would have been excluded from the trial analyses. For example, patients with a second tumour would not be randomised into the trial and would therefore be excluded from the analyses. In addition, the patients recruited into STAMPEDE have high-risk or advanced prostate cancer and hence, selecting for patients with worse outcomes. This may again not provide a representative prostate cancer population and may select for patients with worse outcomes.

The routine data also directly leads to selection bias. For example, more expensive procedures may be coded for, in comparison to less expensive ones which may be excluded. This will have an impact on event detection and therefore model development. Socioeconomic status should make no difference to the quality of coding but as described above, patients receiving private care would not have HES data coding and therefore this may restrict the socioeconomic groups in the analyses.

Ascertainment bias may also be present, for example, in the failure to represent all classes of cases. Where patients on particular trial arms were chosen or patients with particular events (for example, skeletal related events, SREs in chapter 4), this may have introduced bias, by excluding other cohorts. Ascertainment bias may also have been present when performing the time to correct treatment analyses in chapter 5. Codes were used to extract the patient cohort and those without the codes were excluded from the analyses. This would have biased the results, for example, by excluding patients on different pathways, with different times to treatment.

There are also temporal biases; over time the coding schemes have changed and also the motivations for coding. This means that outcomes could have been missed within all the studies but also patients could have been excluded from the analyses. Over time clinical pathways change and also clinician preference, which may have also impacted on the ability for the algorithms to function, for example in the chapter 5 (BladderPath) time to correct treatment study.

Design bias was also present. For example, in chapter 3 (neutropenic event study), there was a potential for partially paired data to be present. Hence, it is possible that the effect of two processes could not be separated. Severe neutropenic events may have been due to repeated exposure to chemotherapy, rather than an individual regimen (discussed further in chapter 3).

Detection bias was also present; events could only be detected if they were present in these data. Hence, in all studies the outcomes may not have been completely impacted by an event of interest (for example, chemotherapy or progression) but actually due to the impact of missing data. In addition, there was often no access to a reference for patients that did not experience an event, due to no known 'truth'. Therefore, it is not possible to verify that the patient did not have an event. Bias is discussed throughout this thesis with regards to the individual studies.

This chapter summarised the shared approaches utilised throughout these studies. However, methods unique to each individual study are further presented within chapter 3, 4 and 5. Firstly, the use of routine data to directly identify outcomes, retrospectively, is presented.

3 CHAPTER THREE: Direct RCT data collection (retrospective model)

3.1 Abstract

3.1.1 Introduction:

Lack of clinical and pathological data in routine data sources, means that only particular outcomes can be identified *directly* in the routine data, whilst others need to be *indirectly* identified. The aim of this chapter was to assess the feasibility of using routine data to *directly* identify events. The example was neutropenic toxicity, a serious adverse event (SAE), in the STAMPEDE trial, to investigate the relationship with the timing of docetaxel administration. This example was chosen because docetaxel was initially licenced for castrate-resistant prostate cancer (CRPC), but more recent evidence showed docetaxel at hormone-sensitive prostate cancer (HSPC) diagnosis improved survival. Higher neutropenic toxicity rates were reported in the HSPC trials, but it is unclear if this was due to the timing of docetaxel or differences in the patient case-mix. This presented an ideal setting to identify if toxicity follow-up information could be feasibly identified in the routine data.

3.1.2 Methods:

The STAMPEDE trial patient data and the routine clinical National Health Service (NHS) data were linked (Hospital Episode Statistics (HES), Systemic Anti-Cancer Therapy Dataset (SACT)). Note review was undertaken, which was linked to the routine data to assess neutropenic admission rates and the feasibility of routine data event detection by hormone-sensitivity (HSPC & CRPC docetaxel) at a single site (Number (N)=44). This was used to develop and validate algorithms to detect neutropenic events across the entire data set. The algorithms were also restricted to detect sepsis-only and sepsis plus neutropenia events (N=3642). Missing HES CRPC chemotherapy regimens were also 'inferred' with HES data alone or enhanced with SACT data (N=1573). The rates of events were calculated across settings. Data quality was assessed throughout the analyses.

3.1.3 Results:

Neutropenic admission events were accurately detected from within the HES data (92%; 48/52; cohort four). However, due to the historic data used, HSPC chemotherapy

regimens were not accurately coded in HES (0% regimens; 0/15; cohort one) or SACT (42% regimens; 83/200; cohort three). Therefore, the model was designed to utilise STAMPEDE data for detection of HSPC regimens. However, the accuracy of detecting CRPC regimens in the routine data was higher than at HSPC. The HES data quality improved over the last five data years (neutropenic events detected in the routine data compared to STAMPEDE trial reported events: 2009, 30% to 2013, 100%). The incidence of neutropenic events by HSPC and CRPC chemotherapy varied by method; for most methods, algorithm-detected neutropenic event rates at CRPC were higher than in the published CRPC trial (TAX-327) but similar to or higher than the algorithm-detected HSPC neutropenic event rates.

3.1.4 Conclusion:

This analysis enabled validated investigation into an outcome not possible with the standard trial data. Clinically, these data suggest that docetaxel administered in a CRPC setting has a similar or higher rate of neutropenic admission events than if administered in a HSPC setting and supports docetaxel use in this HSPC setting. However, due to the lack of clinical disease variables (for example, hormone-sensitivity) and poor coding quality in the routine data, the results of these analyses are hypothesis-generating and not conclusive, thus requiring further investigation.

3.2 Declaration

An abstract describing work contained in this chapter was submitted and accepted to be presented at the European Society for Medical Oncology (ESMO) annual meeting (Barcelona, 30th September 2019). A full manuscript is in preparation.

3.3 Introduction

3.3.1 Clinical trial routine data use

In this thesis, I set out to determine the feasibility of using routine data for clinical trial data collection and analysis. There are trial outcomes that can be *directly* identified in these data and outcomes that cannot be. This is because routine data contains limited clinical variables and therefore methods to *directly* collect outcomes may not be possible, so techniques need to be developed to *indirectly* detect such events (explored in chapter 4). These limited variables include clinical and pathological data, and this limitation is largely due to the fact that many of the data sources are focused on resource use rather than patient outcomes. Hence, analysis of some trial endpoints, such as progression, are often not directly possible. However, if the outcome of interest can be explicitly identified in these data, *direct* data analysis is theoretically possible, and is being explored here. These include events such as attendance at clinic, clinical interventions (surgery, chemotherapy) and diagnoses, which are commonly interrogated *directly* in the literature in non-trial non-oncology settings (43) (42). Therefore, this chapter aims to assess the feasibility of such methods in oncology randomised controlled trials (RCT) to answer a clinically relevant question that was previously not possible using traditional trial data collection techniques.

I set out to identify if events constituting serious adverse events (SAE) (113) could be identified using routine data; SAEs are key events for trial conduct. The overall objective was to investigate the data scope for collecting these uro-oncology outcomes. If this was possible, it could be anticipated that other SAEs may be identifiable within the routine data, however, it is imperative that the quality and completeness of the routine data are validated with respect to each individual outcome. The STAMPEDE trial is aiming to reduce long term clinic-based follow-up for certain historical treatment arms and is scoping the possibility of routine data only follow-up on prospective trial arms. Hence, the aim for this thesis was to investigate the feasibility of such methods.

Anecdotally, new rationale and hypotheses can arise during trial conduct; for example, the recent requirement to assess the rates of sepsis occurring on docetaxel chemotherapy, upon extended follow-up (114). However, if the trial case report forms

(CRF) were not designed to collect these data required to test the new hypothesis, then this analysis is not traditionally possible. Furthermore, it is often not possible to retrospectively request this information directly due to recall bias; loss to follow-up and/or the resources needed to interrogate large sample sizes. These limitations can apply both to patients and to clinical sites. Hence, when new hypotheses arise, often single-site audits are conducted, or comparisons are made across trials, using partial data. Where outcomes cannot be identified through randomisation in a single trial, it is appreciated that *'there are inherent dangers associated with any attempt to directly compare the outcomes observed in independently conducted clinical trials'* (115). This is largely due to differing patient study inclusion criteria and therefore heterogeneous populations, but also differing definitions for outcomes, across studies, make comparisons difficult. Thus, I set out to assess *direct* data analysis techniques using neutropenic toxicity as an example, within a single randomised population.

3.3.2 Neutropenic sepsis

Neutropenic sepsis is a common complication from the anticancer agent chemotherapy (116). Incidences as high as 70-100% have been reported after intensive therapy (117). During chemotherapy treatment, infection fighting neutrophils in the blood can be destroyed (called neutropenia (agranulocytosis)), leading to increased susceptibility to infection, contributing to both sepsis and mortality events. Neutropenic sepsis is a medical emergency and treatment involves intravenous infusion of antibiotics during a hospital admission (118). One study published an overall in-hospital mortality rate of 9.5% after diagnosis of sepsis, increasing to greater than or equal to 21.4% if the patient had more than one comorbidity (additional conditions) (119). The median length of stay was 11.5 days with a median cost per episode of \$8,376 (119). In a United Kingdom (UK) based lung cancer study, the burden of neutropenic sepsis was also investigated (120). The mean length of stay for confirmed neutropenic sepsis was 9.2 days and the mean cost was £3,163 per episode (120).

Therefore, sepsis events account for not only a high mortality rate but also an extensive length of stay with increased resource use and high cost.

3.3.2.1 Pathophysiology

Neutropenia occurs when the number of neutrophils in the blood reduce to defined levels, and either a temperature higher than 38 degrees (febrile neutropenia) or '*signs and symptoms of clinically significant sepsis*' are experienced (121). The body is then susceptible to bacteraemia (infection) and if this leads to a severe inflammatory response, this is defined as neutropenic sepsis (122). The neutrophil nadir usually occurs 10-14 days following chemotherapy treatment and neutrophil levels usually increase again in two to four weeks following the chemotherapy (123). These different patient presentations can often lead to ambiguity in reporting these events.

3.3.2.2 Neutropenic sepsis coding

A coding consultation was launched by NHS Digital (NHSD) regarding the national standard of how to code sepsis events in administrative data sources. The aim was to refine a coding standard for use from the 1st April 2018. This included coding for sepsis, septic shock, severe sepsis and neutropenic sepsis events. It was highlighted that there was large variation in how sepsis events were coded between different hospital trusts, which was reported to potentially be due to variation in the source medical records (124, 125). NHSD proposed that the coding teams worked closely with the clinicians to reduce this variation so that monitoring of sepsis could then be investigated nationally (124, 125). The optimum coding for sepsis events was outlined (125), which contained the codes seen in table 5. However, for events prior to the 1st April 2018, no standardised coding models were available.

Code	Description
A41.X	<ul style="list-style-type: none"> Other sepsis (or the specific sepsis type recorded in the medical record) (124, 125)
R65.1	<ul style="list-style-type: none"> Systemic inflammatory response syndrome of infectious origin with organ failure (only use if the sepsis is documented as severe) (124, 125)
U82.X; U83.X; U84.X	<ul style="list-style-type: none"> U82.X - Resistance to beta-lactam antibiotics; U83.X - Resistance to other antibiotics; U84.X - Resistance to other antimicrobial drugs (only use if the sepsis is resistant to antibiotics or antimicrobial drugs) (124, 125)
D70.X	<ul style="list-style-type: none"> Agranulocytosis (to use in addition to the above codes, to distinguish neutropenic sepsis from severe sepsis) (124, 125)
Further ICD codes (from drugs, medicaments and biological substances causing adverse effects in therapeutic use (Y40-Y59))	<ul style="list-style-type: none"> If documented in the clinical noting that the neutropenia was due to a drug, then an adverse effect code should be used in addition (124, 125)

Table 5: ICD coding that should be utilised post 01/04/18 to document neutropenic sepsis events. Code descriptions taken directly from the coding consultation report (124, 125).

3.3.2.3 Neutropenic events in advanced prostate cancer

Historically, hormone therapy has been the standard-of-care (SOC) for patients with advanced metastatic or high-risk hormone-sensitive prostate cancer (HSPC). However, recently (2015) two large randomised controlled trials (STAMPEDE and CHAARTED) evaluated the use of docetaxel (brand name Taxotere (Sanofi-Aventis)) chemotherapy at diagnosis for HSPC, paradigm shifting treatment in this patient subset. Both STAMPEDE (83) and the CHAARTED trial (126) (ChemoHormonal Therapy Versus Androgen Ablation Randomised Trial for Extensive Disease in Prostate Cancer) reported an increase in survival when docetaxel was administered upfront at diagnosis as first-line therapy for HSPC, rather than as a second-line therapy at disease relapse with castrate-resistant prostate cancer (CRPC). At disease relapse, the prostate cancer that was previously responding to hormone therapy (hormone-sensitive disease, HSPC) can become resistant to the therapy, leading to castration-resistant (CRPC) disease (87). Due to these studies, docetaxel plus hormone therapy is now regarded as the SOC in HSPC (127).

As explained above, chemotherapy is known to lead to neutropenia (128) and potentially severe neutropenic events, defined here as those requiring hospitalisation; for example, severe neutropenia, febrile neutropenia, neutropenic sepsis, infection with neutropenia (further discussed below 3.3.2.4).

The increase in adverse severe neutropenic events was observed in both STAMPEDE and CHARRTED, alongside other studies including local audits (table 6). There was also a raised neutropenic toxicity incidence in the GETUG-AFU15 RCT. GETUG-AFU15 was evaluating ADT alone, or with docetaxel, in HSPC metastatic patients. GETUG-AFU15, however, found a non-significant OS hazard ratio for docetaxel use in the HSPC setting (hazard ratio (HR) Docetaxel + ADT vs. ADT alone: 1.01, 95% confidence interval (CI) 0.75–1.36) (129). In all of these studies the rates of febrile neutropenia were higher than those previously reported for use in the CRPC setting of the TAX-327 (published 2004) trial, at disease relapse (CRPC) (130). The TAX-327 trial was a Sanofi-Aventis pharmaceutical registration study that began recruiting in 2000, providing evidence to license docetaxel in the metastatic CRPC setting (130).

Whilst utilising Docetaxel in a HSPC disease setting (upfront at diagnosis rather than at CRPC for relapsed disease), studies report an increase in serious neutropenic events, when compared to administration at development of CRPC. It has therefore, been proposed that the risk of serious neutropenia-related events may outweigh the benefits of the systemic therapy, if this hypothesis is true.

3.3.2.4 Neutropenic events in the STAMPEDE trial

The STAMPEDE trial was a prime setting to test this hypothesis of increased sepsis at HSPC because of the large-scale patient recruitment. This would ensure a large sample of randomised patients were administered docetaxel in both the HSPC (trial arms C and E) (introduction, figure 3) and the CRPC (patients on all trial arms A-G, at relapse) settings. Hence, a randomised cohort of 3,642 men were available for this analysis.

Neutropenic events are collected in the STAMPEDE trial if they are related to the upfront HSPC docetaxel trial treatment. It is not mandatory to collect additional events due to later non-trial treatments, for example, at relapse with CRPC disease/treatment failure. Therefore, using the standard trial data it was not possible to compare the neutropenic events for HSPC and CRPC patients.

Severe neutropenic events are documented in two distinct STAMPEDE CRFs: 1) Follow-up forms (from trial initiation 2005); these are not reported in real time as require that the patient attends a study follow up visit and 2) Serious Adverse Event (SAE) forms.

Adverse events (AE) and serious adverse events (SAE) are both collected by the trial. AEs are defined as: '*Any untoward medical occurrence in a patient or clinical investigation subject administered a pharmaceutical product and which does not necessarily have to have a causal relationship with this treatment*' (131). In comparison, SAEs are major safety events which are required to be reported to the Medicines and Healthcare Products Regulatory Agency (MHRA) (introduction 1.5.4). SAEs are adverse events that are defined as *serious* (SAEs are therefore a subset of AEs). As per the STAMPEDE protocol, SAEs include AEs that: resulted in death; were life-threatening; required hospitalisation or prolongation of a hospitalisation; resulted in significant disability or incapacity; consists of a congenital anomaly or birth defect or another important medical condition (medical discretion). SAEs are required to be reported within 24 hours of completing the SAE CRF (72) and STAMPEDE also requires that all SAEs should be reported within 30 days after the last trial treatment.

STAMPEDE utilises the Common Terminology Criteria for Adverse Events system (CTCAE) to define trial AE reporting. It enables the standardised allocation of a grade (1–5) to each adverse event experienced (132). The grades are allocated depending on the *severity* of the event. For example, grade one (mild), grade two (moderate), grade three (severe), grade four (life-threatening or urgent intervention required) and grade five (death) (132). Then a subset of these AEs are reported as SAEs if they are defined as *serious* (as per the definition above).

Hence, an event is only deemed to be an SAE if it meets the *seriousness* criteria. This is not the same as the *severity* criteria as used to report AEs. The terms *serious* and *severe* are not synonymous (131). *Severity* is used to describe the intensity of an event (for example, mild, moderate, severe) but this may be to describe an event of *minor medical significance*; for example, a severe headache (131). In comparison, *seriousness*, is describing the outcome of an event, for example, death (131).

Due to this distinction, it is possible that a grade one or two AE could also be an SAE. For example, a mild/moderate event that led to the patient being hospitalised. In addition, not all grade three or four AEs are reported as an SAE. This is because although the definition of *severe* may be met, the event may not be *serious*. Hence, despite the *severity* of grade three and four AEs, it is not possible to classify all grade three and four AEs as *serious* (SAEs). In addition, not all grade five AEs are also reported as SAEs; this is dependent on the timing of the event. If a patient died whilst receiving trial treatment, then this would always be reported as an SAE. Hence, these events would be reported as a grade five AE and an SAE. However, if a patient died greater than 30 days after finishing the trial treatment, then this would not be reported as an SAE. Hence, not all grade five AEs are also reported as SAEs. I am currently aware of a project being undertaken trying to merge these data sources (AEs, SAEs) without duplicating or omitting events; this is not straightforward. However, due to the severity of grade three plus AEs, it is more likely that these events constitute SAEs.

If a low white blood cell count was reported without any sign of infection, this would have been reported as neutropenia and documented in the 'neutrophils' category on the CRF. If there were signs of infection, this should have been reported in the 'febrile neutropenia' category (for example, documenting neutropenic sepsis). Febrile neutropenia should be classified as grade three or higher due to the *severity*. In addition, any neutropenia event (even without infection) that involved hospitalisation should be recorded as a grade three, or higher AE. If the trial treatment led to hospitalisation (for example, severe neutropenic events), this AE should also be reported as an SAE. Due to the potential non-synonymous nature of AEs and SAEs, both sources of data were utilised to try to identify the maximum number of admissions.

Hence, the definitions used to identify STAMPEDE severe neutropenic events for these analyses were either: 1) any febrile neutropenia reports (grade one or higher AEs or SAEs), or 2) neutropenia reports of grade three or higher AEs or SAEs. This should correspond with hospital admissions and therefore it was used as a reference to validate the Hospital Episode Statistic (HES) routine data neutropenic event admissions. The corresponding definitions to identify events from both HES and the clinical noting can be seen in the analyses methods (section 3.11.3).

3.3.2.5 The incidence of neutropenic events from upfront (HSPC) vs. relapse (CRPC) docetaxel

As mentioned above (section 3.3.2.3), an increase in severe neutropenic events were observed across the trials. The rates of febrile neutropenia were reported as follows: STAMPEDE (15%) (83), CHAARTED (6%) (126), and GETUG-AFU15 (8%) (129) (table 6). In all of these studies the rates of febrile neutropenia were higher than those previously reported for use in the metastatic CRPC setting in the TAX-327 trial (3%) (130). The total severe neutropenia rates can be observed in table 6. Table 6 also highlights the interchangeable neutropenic event definitions collected across studies. For example, neutropenia, febrile neutropenia, neutropenic sepsis and infection with neutropenia.

Setting	Author	Study	Grade 3-5 neutropenia (%)	Grade 3-5 febrile neutropenia (%)	Neutropenic sepsis (%)	Infection with neutropenia (%)
HSPC	James (83)	STAMPEDE	12	15	NR	NR
	Sweeney (126)	CHAARTED	12	6	NR	2
	Gravis (129)	GETUG-AFU15	32	8	NR	2
	Mahil (133)	Audit	NR	30*	NR	NR
	Tanguay (134)	Audit	36**	NR	20**	NR
CRPC	Tannock (130)	TAX-327	32**	3*	NR	NR

Table 6: The incidence of neutropenic AEs across trials and audits.

The table was restricted to AEs as SAEs were often not reported by the studies. The types of event, as specified in the studies (neutropenia, febrile neutropenia neutropenic sepsis or infection with neutropenia) have been documented in the table. NR = not reported, * = not specified as grade 3-5, ** = Grade 3 and 4 only.

3.4 Objective

If a patient develops severe neutropenia (neutropenia grade three plus, febrile neutropenia, neutropenic sepsis, infection with neutropenia) post-cycle (chemotherapy), the patient would be admitted to hospital for urgent treatment. I therefore hypothesised that admissions for neutropenic events whilst on chemotherapy could provide us with a proxy marker for severe neutropenic events, due to the lack of outcome reporting in the HES data. Routine data collection for healthcare interactions

provides a unique view of the disease history, which has the potential to be utilised to extract major trial outcomes. The objective of this chapter is to establish the feasibility of using routine data to identify events that were not routinely collected by the STAMPEDE trial.

3.5 Systematic literature review: Introduction

3.5.1 Rationale

As described above, it is a known concern that the largest HSPC docetaxel RCTs presented a higher toxicity rate, than at CRPC. However, since these conclusions were drawn from cross-study comparisons (see table 6), I postulated that a single population study was necessary to compare these outcomes. I first undertook a systematic search to identify publications that examined this controversial sepsis rate, after these pivotal RCTs (table 6) were published to assess the quality of the evidence.

3.5.2 Aims and objectives

The objective of the systematic review was to identify publications and compare the literature surrounding the increased incidence in neutropenic toxicity between HSPC and CRPC docetaxel chemotherapy. The aim was to assess the origin of the current evidence informing clinical practice, since the practice changing RCTs (STAMPEDE, CHARRTED) were published in 2015 (introduction 3.3.2.3), and to assess the quality of the evidence.

3.6 Systematic literature review: Materials and methods

3.6.1 Eligibility criteria

3.6.1.1 Criteria for selecting studies for this review

Types of studies: Any level of clinical evidence, including non-trial studies and response letters to studies.

Types of participants: CRPC patients or advanced metastatic or high-risk HSPC patients.

Types of interventions: Docetaxel chemotherapy, for HSPC or CRPC.

Types of outcomes: Incidence of severe neutropenic events.

The search was not restricted by date or country and can be seen in table 7 (see section 3.6.3 for further details).

3.6.2 Information sources and search strategy

The search was undertaken using two databases: Ovid Medline (since the database began in 1946 to Week 48 (November) 2018) and EMBASE (1980 to Week 48 2018). Articles were identified from the two databases using the search strategy in table 7; search terms were required in the title.

Number	Searches	Results
1	neoplasm*.m_titl.	62584
2	cancer.m_titl.	1872019
3	tumour.m_titl.	88561
4	"malignan*".m_titl.	76782
5	prostate.m_titl.	235970
6	docetaxel.m_titl.	16458
7	taxotere.m_titl.	775
8	sepsis.m_titl.	62244
9	neutropeni*.m_titl.	19940
10	neutropenic sepsis.m_titl.	120
11	febrile neutropenia.m_titl.	3203
12	septic shock.m_titl.	15709
13	1 or 2 or 3 or 4	2082122
14	6 or 7	16796
15	8 or 9 or 10 or 11 or 12	94306
16	5 and 13 and 14 and 15	26
17	remove duplicates from 16	16

Table 7: The Medline search strategy that was used to identify articles for the narrative literature review.

3.6.3 Study selection

After the search, the abstracts and titles were initially screened using the inclusion and exclusion criteria below. The remaining article full texts were then further screened manually for eligibility. The preferred reporting items for systematic reviews and meta-analyses (PRISMA) guidelines (2009 (135)) (figure 7) were adhered to, where possible and appropriate, to document the results from the systematic search (appendix 8.1.1).

Inclusion criteria: Publications regarding the HSPC neutropenic event rate, with or without a CRPC comparison, with or without data analyses, all types of evidence, publications in any language, all types of literature (including grey literature, for example, conference proceedings), literature with an online publication date after the first date of publication (August 2015) of the HSPC docetaxel trials (STAMPEDE; 21/12/15, CHARRTED; 05/08/15) were also included.

Exclusion criteria: Literature where the focus was not on docetaxel-related neutropenic events (search terms were required in the title, but articles were further screened for exclusion), events due to other chemotherapy drugs (for example, cabazitaxel) were excluded. In addition, studies where the risk of neutropenia was predicted or where neutropenia was assessed as a survival marker for chemotherapy, were excluded;

studies investigating the effect of a dose alteration on neutropenic events and conference abstracts for the already included full text were also excluded.

3.6.4 Data collection process

I undertook all of the review; no other investigators were utilised due to lack of resource. All references identified by the search (16 studies) were imported into Endnote Web. Data for the review was collated in Microsoft Excel and a qualitative narrative review was undertaken.

3.6.5 Data items

The following variables were sought; participant setting (advanced metastatic or high-risk, HSPC or CRPC, patients), type of study (any level of clinical evidence, including non-trials and response letters to studies), year of publication (in addition, the e-publication date was identified), intervention (if docetaxel was administered for HSPC or CRPC disease) and outcomes (the incidence of severe neutropenic events). Database, author, title and journal were also extracted but were not discussed during the review.

3.6.6 Risk of bias

A general risk of bias analysis was performed, with the potential sources of bias considered by following the Cochrane handbook guidance (136); *selection*, *performance*, *detection*, *attrition* and *reporting* bias. Due to the heterogeneous types of study and the nature of the type of article (for example, abstract only), a risk of bias assessment could not formally be undertaken. However, the *selection* and *reporting* biases were investigated, where possible, for the three audits identified. The reasons for not identifying risk of bias for the other categories can be seen below in table 8.

Risk	Rationale for the risk of bias assessment
Selection	How patients were allocated to treatments – assessed (important when making cross-study conclusions)
Performance	NR: Blinding of participants is not relevant in this analysis, an RCT was not being conducted – all patients were aware they were being administered chemotherapy and the impacts on the outcomes were therefore assumed to be equal and thus was not investigated
Detection	NR: Blinding of outcome assessors is not relevant in this analysis, an RCT was not being conducted – the rationale is that outcomes are affected by the knowledge of the intervention. All of the assessors were aware of the administration of chemotherapy
Attrition	NR: Missing data – an RCT was not being conducted
Reporting	The reporting of all outcomes – an RCT was not being conducted. However, the comparability of reported outcomes was assessed

Table 8: Rationale for the risk of bias investigation.

NR = not reported

3.6.7 Summary measures

Incidence values of neutropenic events were identified in the literature. The studies did not report 95% confidence intervals (CI) for the incidence values, so these were calculated in R (109), RStudio (110) given the reported data. An asymptotic binomial distribution was assumed. The number of events experienced were also estimated when a percentage number of events experienced in a cohort was provided, but no event number.

3.6.8 Synthesis of results/additional analyses

A meta-analysis was not undertaken due to the heterogeneity in study designs identified.

3.7 Systematic literature review: Results

3.7.1 Study selection

The number of publications identified using each search term can be seen in table 7. 2,082,122 articles were identified in oncology, 235,970 related to prostatic disease, 16,796 related to docetaxel and 94,306 related to neutropenic events. 26 articles were identified that were deemed (by the electronic search) related to neutropenic events occurring due to docetaxel for prostate cancer. One further publication was identified from another source upon a manual search, Google Scholar (137), totalling 27. Once duplicates were removed, 17 papers remained (figure 7).

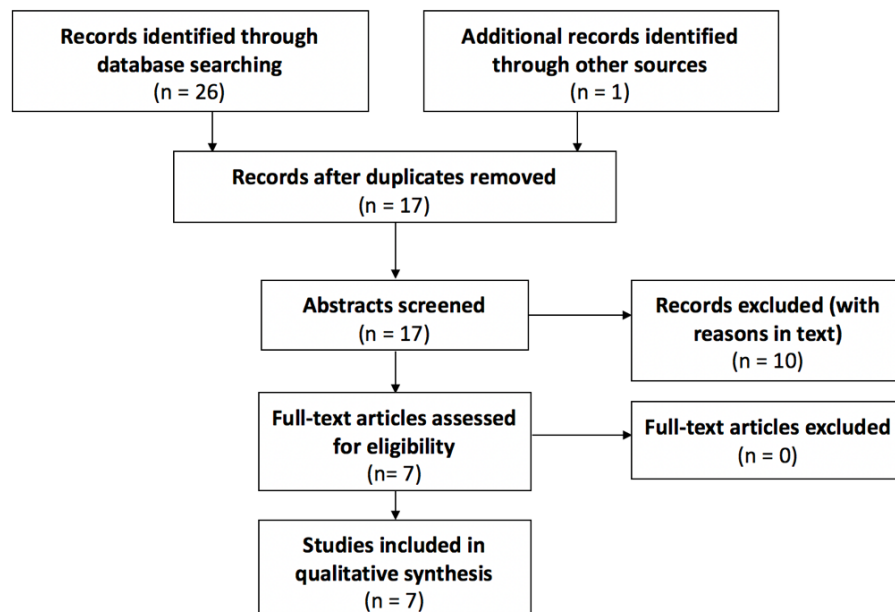


Figure 7: A PRISMA flow diagram showing the stages undertaken to identify the seven papers analysed in the qualitative narrative review.

Ten papers were then manually excluded during abstract screening; those that assessed neutropenia risk factors (N=5); those where chemotherapy was not docetaxel (cabazitaxel) (N=1); those that were presenting a conference abstract for an already excluded full-text (N=1); those that assessed neutropenia as a survival marker (N=2) and those that assessed the effect of a lower dose of docetaxel on febrile neutropenia (N=1). Seven full-text articles remained, which were analysed in the qualitative review.

3.7.2 Summary of the literature identified

The seven publications (five letters, one full-text publication and one conference proceeding) identified in the search are summarised in table 9. Two studies enabled comparison of both HSPC and CRPC patients: Schweizer (138) and Maria (139). Schweizer compared rates within the same population (audit (138)) and Maria, within different populations (meta-analysis (139)). The meta-analysis by Maria examined the toxicity rates of multiple published phase II and III clinical trials and enabled analysis across both settings. Two papers analysed rates in the single hormone setting (HSPC) through single site audits (Tanguay (134), Mahil (133)). As both Tanguay and Mahil analysed data within a single hormone setting, assumptions made of higher HSPC toxicity, than at CRPC, relied upon cross-audit or cross-trial comparisons, as did Maria despite the multiple site nature of a meta-analysis (139). There were also three letters discussing this HSPC toxicity concern (Tsao (140), James (141) and Sydes (114)).

Setting	Author	Year published (online)	Title	Journal	Type
HSPC	Tanguay (134)	2016	High Risk of Neutropenia for Hormone-naïve Prostate Cancer Patients Receiving STAMPEDE-style Upfront Docetaxel Chemotherapy in Usual Clinical Practice.	Clinical Oncology	Audit (single site)
	James (141)	2016	Response to 'High Risk of Neutropenia for Hormone-naïve Prostate Cancer Patients Receiving STAMPEDE-style Upfront Docetaxel Chemotherapy in Usual Clinical Practice', by Tanguay <i>et al.</i>	Clinical Oncology	Tanguay response
	Tsao (140)	2016	Docetaxel for Metastatic Hormone-sensitive Prostate Cancer: Urgent Need to Minimize the Risk of Neutropenic Fever.	European Urology	Letter
	Sydes (114)	2016	Reply to Che-Kai Tsao, Matthew D. Galsky, and William K. Oh's Platinum Opinion. Docetaxel for Metastatic Hormone-Sensitive Prostate Cancer: Urgent Need To Minimize The Risk Of Neutropenic Fever.	European Urology	Tsao response
	Mahil (133)	2016	Febrile Neutropenia Rates in Men Treated with Docetaxel Chemotherapy for Metastatic Hormone-sensitive Prostate Cancer.	Clinical Oncology	Audit (single site)
HSPC & CRPC	Maria (139)	2017	Association of risk of febrile neutropenia (FN) with docetaxel in prostate cancer (PC) patients: A meta-analysis of published phase II-III trials.	Journal of Clinical Oncology (ASCO conference)	Meta-analysis (aggregate)
	Schweizer (138)	2016	Docetaxel-Related Toxicity in Metastatic Hormone-Sensitive and Metastatic Castration-Resistant Prostate Cancer	Medical oncology	Audit (single site)

Table 9: Summary of the search results.

3.7.3 HSPC vs. CRPC neutropenic toxicity

Varying toxicities were documented for HSPC patients across the different studies (table 10). The audits by Mahil (133) and Tanguay (134) suggested that the incidence of neutropenic events was higher in the HSPC setting (table 10) than the CRPC setting (compared to TAX-327) and that the incidence of events may also be higher than previously reported by STAMPEDE (83), GETUG-AFU15 (129) and CHAARTED (126). This was also summarised by Tsao in an opinion piece (140). This was in contrast to the conclusions by Maria (139) (aggregate meta-analysis) and Schweizer (audit) (138), both reporting no significant differences by hormone sensitivity.

Author	Setting	Type	Number patients	Number of severe neutropenic events (HSCP or CRPC) / Number of docetaxel regimens (%)				AE / SAE reported?
				Grade 3-5 neutropenia	Grade 3 -5 febrile neutropenia	Neutropenic sepsis	Infection with neutropenia	
Maria (139)	HSPC	Meta-analysis	X/5,088	NR	NR (12)*	NR	NR	NR
Mahil (133)		Audit	53	NR	16/53 (30)*	NR	NR	NR
Tanguay (134)		Audit	39	14/39 (36)**	NR	8/39 (21)**	NR	NR
Schweizer (138)		Audit	22	7/22 (32)*	2/22 (9)***	NR	NR	NR
James (83)		STAMPEDE (RCT) [^]	550	66/550 (12)	84/550 (15)	NR	NR	AE
Sweeney (126)		CHAARTED (RCT) [^]	390	47/390 (12)	24/390 (6)	NR	9/390 (2)	AE
Gravis (129)		GETUG-AFU15 (RCT) [^]	189	AE: 61/189 (32); SAE: 40/189 (21)*	AE: 15/189 (8); SAE: 6/189 (3)*	NR	AE: 5/189 (3); SAE: 2/189 (1)*	AE + SAE
Schweizer (138)	CRPC	Audit	39	10/39 (26)*	2/39 (5)***	NR	NR	NR
Maria (139)		Meta-analysis	X/5,088	NR	NR (7)*	NR	NR	NR
Tannock (130)		TAX-327 (RCT) [^]	332	105-107/332 (32)**	9-11/332 (3)*	NR	NR	AE

Table 10. HSPC and CRPC neutropenic event (neutropenia, febrile neutropenia, neutropenic sepsis or infection with neutropenia) incidence values identified in the review studies (terminology as reported within each individual study).

[^] = The incidence of the four RCTs (STAMPEDE, CHARTED, GETUG-AFU15 and TAX-327), which are used in the comparisons; NR = Not reported; * = not specified as grade 3-5; ** = grade 3 and 4 only; *** = Defined as neutropenic fever. As TAX-327 did not publish the absolute number of events, this number was estimated from the percentage rate published. The window of recording toxicity events varied across trials. For example, the result reported above in STAMPEDE included events up to 30 days after discontinuation of protocol treatment. The STAMPEDE protocol states, 'All Serious Adverse Events (SAE) and Adverse Events (AE) are reportable from the time of randomisation until 30 days after discontinuation of protocol treatment (refer to Section 1.1.2)' (72). CHARTED did not specify the reporting window, other than that results were reported for where follow-up was available, therefore the entire trial period was potentially used. GETUG-AFU15 reported events in the first 6 months of treatment and TAX-327 collected events that occurred or worsened during treatment, but no further information was provided; X = The number of patients in each group were not published, 5,088 patients were included in the meta-analysis across seven studies but the number of HSPC or CRPC patients was not published.

Tanguay (134) undertook an audit of HSPC docetaxel neutropenia and neutropenic sepsis events (N=39) at a single site. In this audit, grade 3 and 4 neutropenia was experienced in 14/39 (36%) of patients and grade 3 and 4 neutropenic sepsis in 8/39 (21%) of patients (although this value was incorrectly rounded in the published article to 20%) (table 10). In the Mahil (133) single site audit of HSPC febrile neutropenia events (N=53), 16/53 (30%) developed an event. The two following issues were described in the single setting non-trial audits (Mahil, Tanguay), 1) a higher event incidence at HSPC was identified than in the STAMPEDE HSPC trial population and 2) the HSPC incidence was higher than in the CRPC group, when compared to TAX-327 (figure 8).

In contrast, Schweizer (138) conducted an audit of non-trial (N=61) patients at a single-site. 2/22 (9%) developed neutropenic fever in the HSPC setting compared to 2/39 (5%) in the CRPC setting but still concluded this *'did not demonstrate that docetaxel is more toxic in the mHSPC (m = metastatic) setting compared to the mCRPC setting'* due to a non-significant p-value (p=0.95). The aggregate meta-analysis by Maria (139) included seven RCT studies (total number of patients across groups: N=5,088); 6.6% of patients that received docetaxel in the CRPC setting developed febrile neutropenia, compared to 12.4% in the HSPC setting (the number of patients in each hormone state were not documented). Despite this, there was also no statistically significant difference (p=0.7) in the febrile neutropenia rate by hormone-sensitivity (HSPC 12.4%; CRPC 6.6%). The exact model used to generate the incidence and P-values were not reported. However, the risk ratio calculations were conducted using the Mantel-Haenszel method with a fixed-effects model.

When comparing the incidence values for febrile neutropenia or neutropenic sepsis, from the single hormone-state studies (Mahil, and Tanguay) to the CRPC TAX-327 trial, an increased toxicity in the HSPC setting was seen (30% Mahil, 21% Tanguay, 3% TAX-327). The meta-analysis by Maria and the audit by Schweizer, also identified an increase in febrile neutropenia events compared to TAX-327 (12% and 9% respectively). However, when comparing both HSPC and CRPC groups in the Schweizer and Maria studies, no significant differences between groups were identified (figure 8).

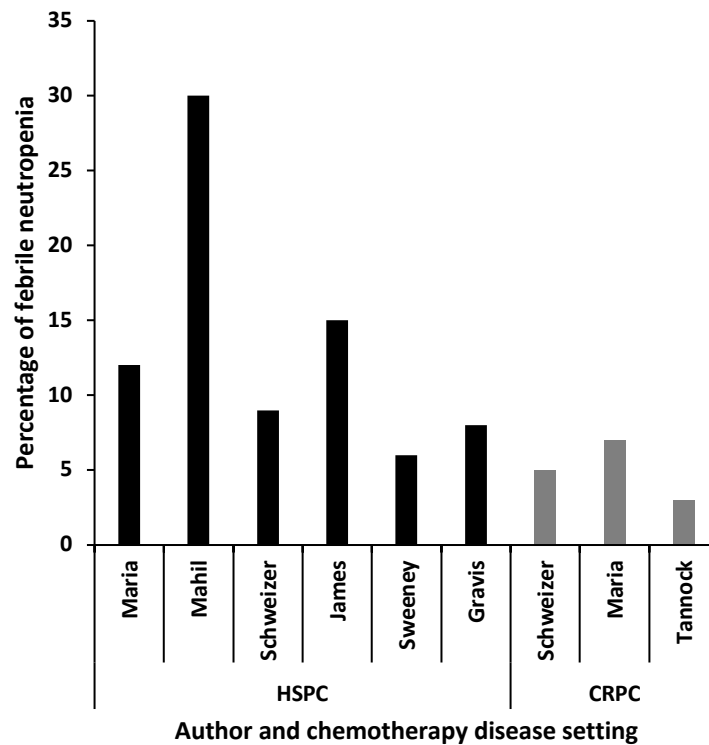


Figure 8: Across study reported rates of febrile neutropenia.
One study (Schweizer) used a different terminology for febrile neutropenia (neutropenic fever).

Confidence intervals were estimated for the event incidence rates, as these were not reported (table 11).

Setting	Author	Type	Events / N patients (% events)	95% CI
HSPC	Mahil (133)	Audit	16/53 (30)	0.18-0.43 (febrile neutropenia)
	Tanguay (134)	Audit	8/39 (21)	0.08-0.33 (neutropenic sepsis)
	Schweizer (138)	Audit	2/22 (9)	0.00-0.21 (neutropenic fever)
	STAMPEDE*	RCT	84/550 (15)	0.12-0.18 (febrile neutropenia)
	CHARRTED*	RCT	24/390 (6)	0.04-0.09 (febrile neutropenia)
	GETUG-AFU15*	RCT	15/189 (8)	0.04-0.12 (febrile neutropenia)
CRPC	Schweizer (138)	Audit	2/39 (5)	0.00-0.12 (neutropenic fever)
	TAX-327*	RCT	9-11/332 (3)	0.01-0.04/0.05 (febrile neutropenia)

Table 11: Estimated confidence intervals calculated from the audit AE rates, compared to published RCT incidence values (*).

In the TAX-327 RCT, 3% of 332 developed febrile neutropenia, it was therefore estimated that 9-11 events were experienced (95% CI: 0.01 – 0.04/0.05) (table 11). In the STAMPEDE RCT, there was a 15% febrile neutropenia rate (84/550, 95% CI: 0.12 –

0.18) (table 11). In the CHAARTED RCT, 6% experienced febrile neutropenia (grade 3 or 4) (24/390, 95% CI: 0.04 – 0.09) and in GETUG-AFU15, 7% (14/189) experienced a grade 3-4 febrile neutropenia event and <1% (1/189) experienced a grade 5 event, therefore 8% (15/189) experienced a grade 3-5 event (95% CI: 0.04 - 0.12) (table 11). The confidence intervals for the audits can also be seen in table 11, ranging from 0.00 – 0.43 in the HSPC setting (three studies) and 0.00 to 0.12 in the CRPC setting (single study).

The seven individual studies (figure 8) were analysed to investigate explanations for these increased HSPC toxicity concerns, including: heterogeneous populations (cross-trial/audit comparisons), translation of results to a real-world setting, sample size and interchangeable definitions, discussed below.

3.8 Systematic literature review: Discussion

No randomised, or large-scale non-randomised analyses (the only non-randomised comparison study was by Schweizer, N=61) or comparisons were identified that compared HSPC and CRPC neutropenic events within the same population. In the absence of this, conclusions were drawn from making simple comparisons, which may have led to reporting bias and invalid conclusions being drawn. An increased toxicity in the HSPC setting was concluded by Mahil and Tanguay, whilst making cross-study comparisons (due to conducting single hormone-setting analysis) to the CRPC TAX-327 trial. It is difficult to generalise these conclusions drawn from cross-study comparisons, due to heterogeneity in the patients recruited, for example, with a difference in baseline inclusion criteria. This may mean that some cohorts have a fitness advantage, biasing the outcome.

In contrast, when comparing HSPC and CRPC febrile neutropenia rates within the same population (Schweizer (138)), there was no statistically significant difference between the groups, despite a higher rate in the HSPC group. All of these audits had small cohort numbers and were run at single-sites; Mahil (N=53), Tanguay (N=39) and Schweizer (N=61). In Maria's meta-analysis, there was an increase observed in the HSPC group, and again, there was no statistical difference between the groups. Although the meta-analysis was based upon cross-trial comparisons, this is a better level of evidence than comparing the results of single audit studies, due to higher power, with an overall larger cohort size. When the population was homogenised, by either increasing sample size by including more trials (Maria), or investigating patients from within the same site (Schweizer), no significant difference was identified between the groups. Hence, single cross-study comparisons may have been leading to incorrect conclusions.

Furthermore, the studies did not report confidence intervals (CI) for the incidence values. However, where possible, I estimated the 95% CIs, which are shown in table 11. In the HSPC setting, wider confidence intervals can be seen (Mahil (133), Tanguay (134) and Schweizer (138)) due to the smaller sample sizes (table 11). The lower limits in the HSPC group of 0.08 (Tanguay (134)) and 0.00 (Schweizer (138)), suggest that the HSPC incidence of severe neutropenic events could have been very low; this would have provided an alternative conclusion of a lower HSPC incidence than STAMPEDE and even

TAX-327 (in the Schweizer study). The TAX-327 intervals also overlapped with the HSPC trials CHAARTED and GETUG-AFU15; giving further potential evidence towards less distinct rates by hormone-sensitivity. The above HSPC toxicity rates should therefore be interpreted with caution due to these overlapping confidence intervals, across studies.

Where there is ambiguity in the documentation and classification of events (124, 125) (such as neutropenic events), it is difficult to make cross-study comparisons. For example, the terms neutropenic sepsis and febrile neutropenia are often used interchangeably (122). This lack of standardised criteria can be seen across the studies above (table 10), resulting in difficult cross-study comparisons. The three single-site audits (Mahil (133), Tanguay (134) and Schweizer (138)) reported different types of neutropenic events (table 10); Mahil (133) reported febrile neutropenia but the grade was not specified, Tanguay (134) reported neutropenia and neutropenic sepsis but grade 3 and 4 only and Schweizer (138) reported neutropenic fever but did not specify the grade. Therefore, comparison of febrile neutropenia rates across audits was not possible. Tanguay reported HSPC neutropenia and neutropenic sepsis rates and compared these to the STAMPEDE rates. However, although STAMPEDE reported a neutropenia rate, STAMPEDE did not publish a neutropenic sepsis rate and instead reported febrile neutropenia. Tanguay also wrongly reported the published STAMPEDE rate of febrile neutropenia and as previously stated, incorrectly rounded the neutropenic sepsis audit rate. Maria (139) quoted a febrile neutropenia rate but did not specify the grade, nor was it possible to determine how the included studies graded events.

In the HSPC RCTs, STAMPEDE, CHAARTED and GETUG-AFU15 all published grade 3-5 neutropenia rates, but only grade 3-4 were reported in TAX-327 (CRPC). Grade 3-5 febrile neutropenia was also reported in STAMPEDE, CHAARTED and GETUG-AFU15 but the grade was not specified in TAX-327. Neutropenic sepsis rates were not reported by the RCTs, although neutropenic sepsis is the term often used. GETUG-AFU15 was also the only study to report both AE and the subset of SAEs, therefore it was not possible to compare SAEs across the trials. It cannot always be assumed that all grade 3+ events resulted in hospitalisation. Therefore, not all grade 3+ AEs equal SAEs; this can be seen in the greater number of AEs than SAEs in GETUG-AFU15, despite the AEs being defined

as grade 3+ (see section 3.3.2.4 for further details). As reported by Sydes (114), these varying toxicity criteria and classification of events make audit-trial comparisons and cross-trial comparisons challenging and '*greater standardisation in reporting is required*' (114).

There were also further considerations making conclusions drawn from the published literature challenging; it was often not reported whether drugs had been administered in combination with a drug called granulocyte colony-stimulating factor (G-CSF) (to help reduce the damage to the white blood cells), this may have had incidence implications. Furthermore, the rates of febrile neutropenia reported in the audits will have been identified from clinical noting, this may not have documented the neutrophil count or fever and may have led to an overestimation in cases. Some events may have also been missing if they occurred at other hospitals (142). Some events may also have been checked (screened) for more frequently; therefore, more events would potentially have been reported. For example, the TAX-327 licencing trial regularly checked for neutropenia using laboratory abnormalities. In contrast to STAMPEDE, when docetaxel was an established treatment, laboratory screening was not routinely undertaken, hence a higher reporting rate of neutropenia was expected in TAX-327.

Patient selection can also lead to a risk of bias. In small single-site audits, the results may reflect local clinical practice, for example, patient selection for treatment, rather than true drug effects. Patients may have been selected based upon fitness and fitter patients may be less at risk of events. No patient characteristic tables including age, Charlson comorbidity scores (143) or disease stage were included in the audits/meta-analyses to enable comparison.

3.9 Systematic literature review: Conclusion

Much of the evidence of higher toxicity burden when utilising docetaxel for HSPC arose from making cross-trial comparisons which are inherently challenging (144).

Furthermore, these comparisons were made in the absence of randomisation; randomisation is the 'gold standard' technique to control differences in patient characteristics and other factors, that may otherwise bias conclusions when making cross-study comparisons. I explored possible explanations for these different neutropenia event rates, including, small audit sample sizes effecting significance (133, 134), overlapping confidence intervals and interchangeable definitions (134, 140, 141).

Due to an absence of any randomised studies comparing rates, there is an unmet clinical need to resolve this toxicity uncertainty. Hence, this is an ideal setting to investigate if routine data could be used to perform analysis not previously possible with standard trial designs. This chapter outlines a method to utilise routine data to enable direct comparison in the same patient population. Due to the broad eligibility criteria of the STAMPEDE trial (72) and the vast numbers of patients recruited, I believed that valid within-trial comparisons of neutropenic event rates could be undertaken, whilst assessing the feasibility of using routine data to perform such analyses.

3.10 Use of routine data to evaluate severe neutropenic events in patients administered docetaxel for HSPC and CRPC: Rationale

Routine healthcare data can contain limited variables (clinical/pathological details), necessitating the development of *indirect* models to enable outcome analysis (explored in chapter 4). However, if the routine data contains the variables required to answer a research question, then these data can be interrogated *directly* (introduction 3.3.1). In order to assess the feasibility of using these data *directly* for oncology clinical trial outcome reporting, neutropenic AEs were investigated within the STAMPEDE trial, as an example. The use of routine data was evaluated whilst comparing severe neutropenic events upon the administration of docetaxel chemotherapy for HSPC and CRPC.

3.11 Materials and methods

3.11.1 Approvals

Details of the approvals required for this study can be found in chapter 2, section 2.4. For example, STAMPEDE trial consent for routine data access; STAMPEDE ethics approval (hence, this study is consequently an ancillary study within the STAMPEDE trial, as detailed in section 17.4 of the protocol) (72); registration for the University Hospitals Birmingham, Queen Elizabeth Hospital (UHB QEH) audit and applications to the Medical Research Council, Clinical Trials Unit at University College London (MRC CTU at UCL) for STAMPEDE database access. As detailed in section 2.4.1, routine HES data for the single site UHB QEH were previously provided upon application to the Health and Social Care Information Centre (HSCIC, now NHSD) (27). For multiple-site data analyses I applied to Public Health England (PHE) National Cancer Registration and Analysis Service (NCRAS) for multiple dataset access. The HES (admission data set) and Systemic Anti-Cancer Therapy Dataset (SACT) (cycle and regimen datasets used) were utilised during these analyses (see table 12 and figure 9) (52).

Dataset	Data start	Data end
HES admission data	14/11/2005	31/01/2018
SACT cycle data	09/04/2010	31/12/2017
SACT regimen data	23/06/2009	31/12/2017

Table 12: The earliest and latest interaction events identified in the HES and SACT datasets, provided by PHE.

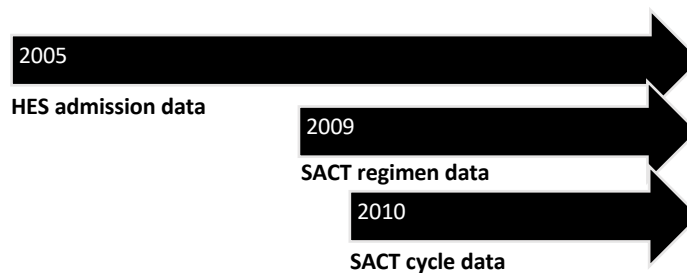


Figure 9: The earliest and latest interaction events identified in the HES and SACT datasets. The figure graphically displays the information shown in table 12.

A description of the cohorts and the rationale can be seen below in section 3.11.2.

3.11.2 Cohorts

Four distinct cohorts were analysed, and participants may have fit in multiple cohorts (table 13, table 14, figure 10). The overall rationale for inclusion was as follows: all participants were enrolled in the STAMPEDE study in England. The cohorts chosen were dictated by the trial arms, the date of randomisation with regards to the availability of the routine data and further routine data specifics (data available and linkage successful) (figure 11).

As STAMPEDE uses a Multi-Arm-Multi-Stage (MAMS) trial design (section 1.1 and 1.5.1), patients were analysed across multiple treatment arms (A-G). See the introduction 1.5.1 for a description of the treatment arms and the trial protocol for further treatment arm details (72).

Patients were required to have consented for routine data access. Participants included in the analyses were randomised between 15/11/2005 and 17/01/2014 (table 13). However, for the cohort three analyses, only patients randomised after 01/04/12 could be included due to the availability of the SACT data, which in turn reduced the number of trial arms which could be analysed (i.e. D or F patients were excluded) (figure 11) (figure 12).

Cohort	One	Two	Three	Four
Number of patients	44	113 (69 unique vs. cohort 1)	1,573 (1,548 unique vs. cohort two)	3,642 (2,069 unique vs. cohort three)
Arm	A-F	A-G	A, B, C, E, G	A-G
Randomised	23/05/06-22/02/13	28/02/06-20/12/13	01/04/12-17/01/14	15/11/05-17/01/14
Site	Single-site (UHB QEH)	Single-site (UHB QEH)	Multi-site (England)	Multi-site (England)

Table 13: The four cohorts used in the analyses.

Cohort one and two were single site studies, however, cohort three and four were large multi-site analyses (table 13). The number in each cohort can be seen in table 13 and the relationship between cohorts in table 14, figure 10 and 11.

Cohort	Cohort				Total
	1	2	3	4	
1	44	44	3	41	44
2	44	113	25	106	113
3	3	25	1573	1573	1573
4	41	106	1573	3642	3642
Total	44	113	1573	3642	-

Table 14: Cross-tabulation of STAMPEDE participants and membership in each cohort.
For example, cohort two and three shared 25 patients and cohort three and four shared 1,573.

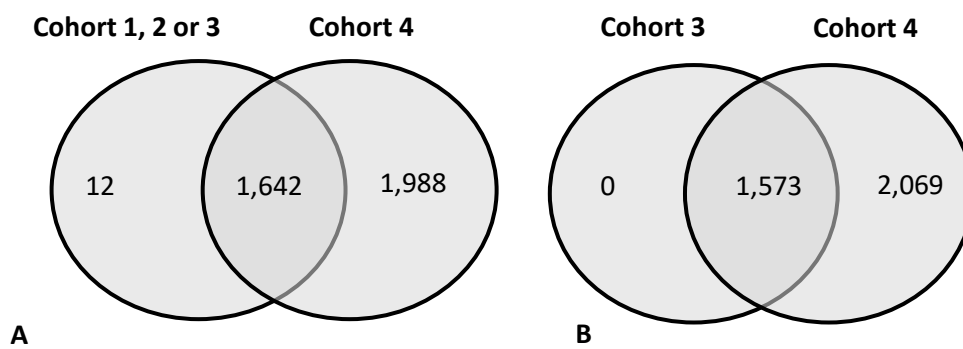


Figure 10: Venn diagrams illustrating the overlap between study cohorts.
A) The patient overlap between the largest cohorts (four) and the other cohorts (one – three). B) The patient overlap between cohort three and four.

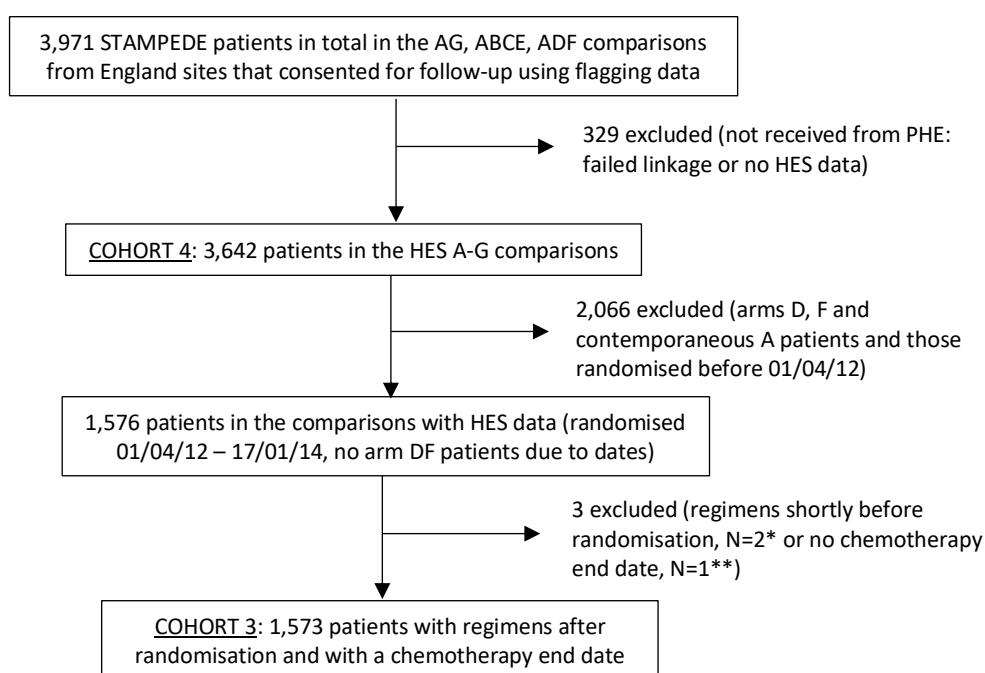


Figure 11: Patients included in the large-scale analyses, cohorts three and four.

** = Docetaxel HSPC regimens were identified using the STAMPEDE database for all algorithms in the cohort 3 and 4 analyses. However, additional data available in the SACT dataset identified extra chemotherapy regimens shortly before randomisation, these patients were excluded (N=2), due to the potential for confounded classification of HSPC treatment. ** = Where it was not possible to infer a chemotherapy end date, the patients were excluded (N=1). It was not possible to classify the hormone-setting of these events by date using the rules, due to the missing data.*

For the multi-site studies, the 3,971 patient cohort comprised of all eligible STAMPEDE England arm A-G patients that had been identified as consenting to data linkage at the time of the final analyses (figure 11). After linkage to the routine data by the provider and after exclusions due to date and event timeliness (for example, after randomisation), 1,573 and 3,642 patients were available for cohort three (SACT and HES) (figure 12) and four (HES-only) analyses, respectively (figure 11) (figure 12).

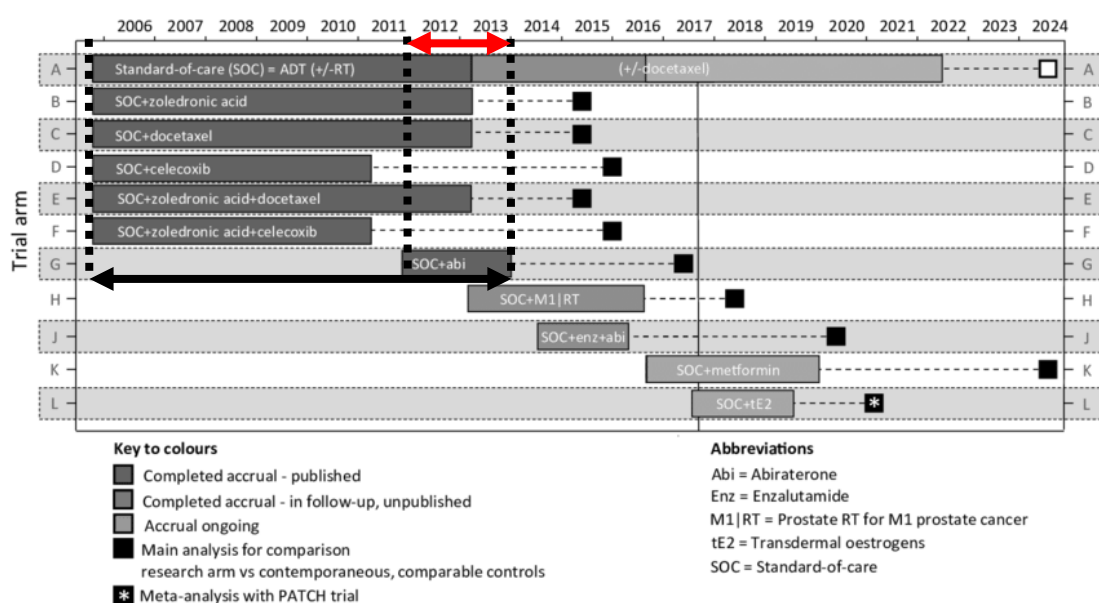


Figure 12: Cohorts analysed from the STAMPEDE trial.

Red arrow: For cohort three analyses, only A, B, C, E and G were recruiting whilst the SACT data were available (patients randomised between 01/04/12 – 17/01/14). Black arrow: The other cohort analyses (one, two and four) included patients from all arms, as the HES data were not restricted by date; therefore all patients could be captured and hence analysis included patients on arm A-G. Figure adapted from the STAMPEDE trial protocol (72).

3.11.3 Data sources and events of interest

Data were extracted for the cohorts from three separate sources (post-STAMPEDE randomisation): 1) STAMPEDE trial database, 2) routine data records (HES and SACT datasets) and 3) reference clinical noting (single site UHB QEH) (see section 2.3). Events could be present in one or all sources (figure 13).

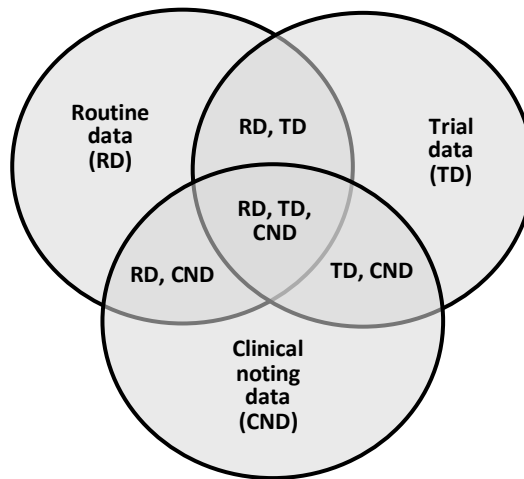


Figure 13: Figure to show the three separate sources of data and their intersections.

RD = routine data (HES, SACT); TD = trial data (STAMPEDE trial); CND = clinical noting data (single site).

The elements extracted from the three datasets can be seen in table 15. The STAMPEDE database records that were extracted included: treatment allocation, neutropenic events plus when the event occurred (weeks post-randomisation) and randomisation date. Routine data and clinical noting records of interest were extracted including: neutropenic admission events (see table 16 for routine data codes) and chemotherapy regimens (see table 17 for routine data codes), plus date.

	Extracted data element	Clinical noting or STAMPEDE extracted	Routine data extracted
1	HSPC docetaxel given?	✓	✗
2	Later chemotherapy given upon relapse? [^] (proxy for CRPC chemotherapy)	✓	✓
3	Chemotherapy drug	✓	✓*
4	Estimated week commenced	✓	✓*
5	Neutropenic events (severe – not including blood only changes)	✓	✓
6	The number of weeks post randomisation of the event	✓	✓
7	Severe neutropenic events due to HSPC or CRPC chemotherapy	✓	✓
8	Further notes on the event	✓	✗

Table 15: The data elements extracted from the clinical noting, STAMPEDE data and the routine data.

* = identifiable in the SACT data only, not HES. ^ = confirmed via note review. However, when the events were extracted from the routine data, in the absence of a reference, these events were inferred using a validated proxy.

The clinical records were considered the reference standard and were analysed via a case series note review to identify the event of interest: admission for suspected neutropenic event. This was defined as: *admission to hospital for suspected neutropenic sepsis/febrile neutropenia/severe neutropenic event*, plus a requirement to be *on a chemotherapy regimen* at the time of the event. Biochemical neutrophil count was not investigated during the note review.

Code	Name
OPCS coding	
X903	Neutropenia drugs band 1
ICD coding	
D70X*	Agranulocytosis
A400	Sepsis due to streptococcus, group A
A401	Sepsis due to streptococcus, group B
A402	Sepsis due to streptococcus, group D
A403	Sepsis due to Streptococcus pneumoniae
A408	Other streptococcal sepsis
A409	Streptococcal sepsis, unspecified
A410*	Sepsis due to Staphylococcus aureus
A411*	Sepsis due to other specified staphylococcus
A412*	Sepsis due to unspecified staphylococcus
A413*	Sepsis due to Haemophilus influenzae
A414*	Sepsis due to anaerobes
A415*	Sepsis due to other Gram-negative organisms
A418*	Other specified sepsis
A419*	Sepsis, unspecified
R572	Septic shock
R579	Shock, unspecified
R651*	Systemic inflammatory response syndrome of infectious origin with organ failure

Table 16: Codes identified in the routine data by the algorithms to identify neutropenic admissions.

* = codes referenced in the NHS coding consultation detailed in the introduction for this chapter (section 3.3.2.2). Bold = codes utilised for the restricted 'sepsis/septic shock' only analysis.

Chemotherapy procurement	
X701	Procurement of drugs for chemotherapy for neoplasm for regimens in Band 1
X702	Procurement of drugs for chemotherapy for neoplasm for regimens in Band 2
X703	Procurement of drugs for chemotherapy for neoplasm for regimens in Band 3
X704	Procurement of drugs for chemotherapy for neoplasm for regimens in Band 4
X705	Procurement of drugs for chemotherapy for neoplasm for regimens in Band 5
X708	Other specified procurement of drugs for chemotherapy for neoplasm in Bands 1-5
X709	Unspecified procurement of drugs for chemotherapy for neoplasm in Bands 1-5
X711	Procurement of drugs for chemotherapy for neoplasm for regimens in Band 6
X712	Procurement of drugs for chemotherapy for neoplasm for regimens in Band 7
X713	Procurement of drugs for chemotherapy for neoplasm for regimens in Band 8
X714	Procurement of drugs for chemotherapy for neoplasm for regimens in Band 9
X715	Procurement of drugs for chemotherapy for neoplasm for regimens in Band 10
X718	Other specified procurement of drugs for chemotherapy for neoplasm in Bands 6-10
X719	Unspecified procurement of drugs for chemotherapy for neoplasm in Bands 6-10
Chemotherapy delivery	
X352	Intravenous chemotherapy
X373*	Intramuscular chemotherapy
X384*	Subcutaneous chemotherapy
X721	Delivery of complex chemotherapy for neoplasm including prolonged infusional treatment at first attendance
X722	Delivery of complex parenteral chemotherapy for neoplasm at first attendance
X723	Delivery of simple parenteral chemotherapy for neoplasm at first attendance
X724	Delivery of subsequent element of cycle of chemotherapy for neoplasm
X728	Other specified delivery of chemotherapy for neoplasm
X729	Unspecified delivery of chemotherapy for neoplasm
X731*	Delivery of exclusively oral chemotherapy for neoplasm
X738*	Other specified delivery of oral chemotherapy for neoplasm
X739*	Unspecified delivery of oral chemotherapy for neoplasm
X748	Other specified other chemotherapy drugs
X749	Unspecified other chemotherapy drugs
Y123*	Electrochemotherapy to lesion of organ NOC (not otherwise classifiable)

Table 17: OPCS codes identified by the HES routine data algorithms to identify chemotherapy regimens.

Nota bene: The SACT + HES algorithm used detailed drug data to identify chemotherapy regimens (SACT acquired) and not the above routine data codes. * = Broader codes included to detect miscoded chemotherapy events.

Chemotherapy was identified in the clinical noting. It was either determined to have been given upfront in STAMPEDE (classified as administered for HSPC), or not, which was therefore classified as being given at relapse despite hormone therapy (72, 145) (classified as administered for CRPC).

Chemotherapy regimens were also identified in the routine data and classified as being administered for the two groups: 1) HSPC or 2) CRPC. Due to the extent of the missing routine data HSPC regimens, these regimens were identified using the STAMPEDE database; if the patient was randomised to trial arm C or E (see figure 12), these patients were classified as having docetaxel for HSPC. Due to the lack of clinical disease variables

in the routine data documenting hormone settings, the timing of administration was used as a proxy for identifying CRPC docetaxel regimens in the routine data. Hence, when the algorithm was using routine data to define hormone settings, these were not defined by a trial documented response to treatment but by a series of assumptions.

It is possible that a patient could have had chemotherapy in both HSPC and CRPC settings. Neutropenic admission events were also identified, and further rules were created to classify whether the routine data-derived neutropenic event occurred due to HSPC or CRPC chemotherapy, or if unrelated. It is possible that a patient had chemotherapy in both HSPC and CRPC settings, hence, it is possible that a patient could have had a neutropenic event in both settings. Therefore, partially paired data are possible (146). However, all patients were included to reflect the reality of clinical practice (see discussion 3.13). Data were treated as independent events and no further statistical tests were undertaken to account for this.

Once a neutropenic event was identified, it was presumed to be associated with chemotherapy if it occurred within four weeks of the last cycle administration. This is due to the neutrophil nadir which was discussed in section 3.3.2.1 (123). Neutropenic events outside of the four-week time frame were assumed to be either due to CRPC chemotherapy or unrelated to chemotherapy. These events were assessed to identify if they were CRPC-related using a series of rules (as described in section 3.11.7.1). If the event was determined to be unrelated to chemotherapy, they were excluded from the analyses.

3.11.4 Data censor

Dates of available follow-up data for each cohort differed by data source and are shown in table 18.

- **Cohort one, two and four:** Data for each participant were analysed from the date of STAMPEDE randomisation (table 18) and censored at the last available HES episode (HES censor), even if data from other sources were available (table 18).

- **Cohort three:** Data were analysed from the first patient randomised (table 18) after SACT data began being collected (01/04/12) and censored at the SACT data freeze (31/12/17) (table 18). Although SACT collection began prospectively in April 2012, collection at sites became mandatory from July 2015. Retrospectively input data prior to prospective data collection in April 2012 were excluded from the analysis, hence, data analysis began on the 01/04/12. Due to performing analyses on restricted published STAMPEDE cohorts (A-G), no patients randomised past 17/01/14 were available for this study. Therefore, all events were prior to mandatory data collection, with clear accuracy implications.

Cohort	Start/censor	STAMPEDE randomisation date	HES admission	SACT cycle & regimen
Cohort 1	Start analysis	23/05/06	23/05/06	NA
	Censor	22/02/13	30/07/16	NA
Cohort 2	Start analysis	28/02/06	28/02/06	NA
	Censor	20/12/13	31/03/17	NA
Cohort 3	Start analysis	01/04/12	01/04/12	01/04/12
	Censor	17/01/14	31/12/17	31/12/17
Cohort 4	Start analysis	15/11/05	*14/11/05	NA
	Censor	17/01/14	31/01/18	NA

Table 18: Censor dates used for the analyses.

* = Data were provided six months prior to randomisation; therefore, it was possible to find events prior to randomisation, hence, an admission was reported one day prior to randomisation.

3.11.5 Routine data linkage

The routine HES and SACT data were linked to the STAMPEDE trial data by PHE NCRAS (as described in 2.4.1). The trial to routine data linkage rates were calculated as the number of patients that were linked to the routine data sources (HES and SACT), as a percentage of the total.

3.11.6 Data preparation

Data preparation was required to extract meaningful data from the routine data resources. The HES data required for the analyses were present within one data table, the HES admitted patient care (APC) (table 19).

HES APC
STAMPEDEID
ADMIDATE
DIAG (all fields) (ICD codes)
OPERTN (all fields) (OPCS codes)

Table 19: Fields utilised from the HES data table for analysis.

The SACT data required were between two tables, the SACT REGIMEN and CYCLE table. Therefore, the SACT data had to be processed in order to make a single analysable dataset. The SACT REGIMEN and CYCLE table were merged (by the unique variable MERGED_REGIMEN_ID_PS) to create a data table of required fields (table 20) containing a start and an end date of treatment.

REGIMEN table	CYCLE table
STAMPEDEID	STAMPEDEID
MERGED_REGIMEN_ID_PS	MERGED_REGIMEN_ID_PS
ANALYSIS_GROUP	START_DATE_OF_CYCLE
START_DATE_OF_REGIMEN	
DATE_OF_FINAL_TREATMENT	

Table 20: Fields utilised from the SACT data tables for analysis.

When not available in the REGIMEN table, the dates of final treatment were identified as the *last* cycle date (variable: START_DATE_OF_CYCLE) from the CYCLE table. The dates of final treatment (variable: DATE_OF_FINAL_TREATMENT) were often missing and sometimes the values were prior to the *last* cycle dates in the CYCLE table. Therefore, a rule was created: if the date of final treatment (variable: DATE_OF_FINAL_TREATMENT) was prior to the *last* (final) cycle date (variable: START_DATE_OF_CYCLE), then substitute the date of the final treatment (variable: DATE_OF_FINAL_TREATMENT) with the *last* cycle date (variable: START_DATE_OF_CYCLE).

3.11.7 Analytical methods

3.11.7.1 The algorithms

Algorithms were developed in the software R v3.3.2 (109) using RStudio v1.0.136 (110) to identify the events of interest from the routine data (chemotherapy (table 17) and neutropenic admissions (table 16)) (code available on request). Incidence values could then be calculated as the *proportion of patients experiencing a neutropenic admission whilst on HSPC or CRPC chemotherapy*.

Rules/assumptions

The algorithms were based upon a series of rules/assumptions (table 21) to extract the relevant Office of Population Censuses and Surveys Classification of Interventions and Procedures (OPCS) procedure codes (33) and International Statistical Classification of Diseases (ICD) diagnosis (32) codes to identify the neutropenic and chemotherapy events. The neutropenic events then had to be classified as occurring in the HSPC or CRPC setting or if were unrelated to chemotherapy (figure 14). An algorithm was developed to use HES data alone, but in addition missing CRPC regimens were inferred with HES or supplemented using SACT data.

The main classifications were to:

- Identify and classify *chemotherapy* regimens as being administered for HSPC or CRPC. STAMPEDE data (treatment arm) were linked to the HES data to ensure maximum HSPC chemotherapy regimen detection. The CRPC regimens were to be identified in the routine data alone (both HES and SACT) as they are not required to be documented in the trial data.
- Identify *neutropenic* events and classify them as being HSPC or CRPC chemotherapy-related or unrelated. Admission events were identified using HES coding only.

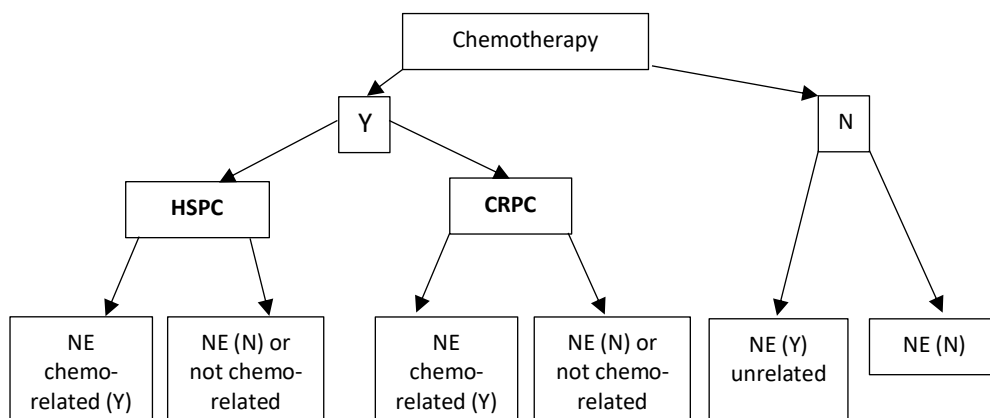


Figure 14: The potential classification of events from the administration of chemotherapy.

NE = neutropenic event; Y = yes and event occurred; N = no event occurred.

In order to classify these events, nine main rules were developed (table 21) and applied sequentially for the routine data algorithms.

#	Rule	Rationale summary
Correctly identify & classify HSPC and CRPC chemotherapy		
1	Identify HSPC regimens from the STAMPEDE database	Due to merging the routine data with the STAMPEDE database, it was known if the patient was administered docetaxel for HSPC (arm C, E). Those that had a HSPC docetaxel arm documented in STAMPEDE were identified as being administered docetaxel for HSPC, those not randomised to a chemotherapy containing arm were not (A, B, D, F).
2	Identify CRPC regimens from the HES data (where a HSPC regimen was not identified)	If chemotherapy was identified in the routine data but the patient was not administered HSPC chemotherapy in STAMPEDE, it was implied that the regimen was for CRPC. This may occur shortly after randomisation, as recurrent disease may occur very quickly if randomised to arm A (hormone therapy); initially assume this was relapse CRPC treatment.
3	Identify CRPC regimens from the HES data (where a HSPC regimen was also identified)	Where both HSPC and CRPC regimens were identified (from STAMPEDE and routine data), the event needed to be classified. A 21-week cut off for the last chemotherapy event was used. Where additional data were available in SACT for the start date of treatment, an additional level of integrity could be adopted, as further classification could be undertaken.
Identify & classify neutropenic events as occurring from HSPC or CRPC regimens, or unrelated to chemotherapy		
4	<i>Inferal rule</i> - Inferring missing HES CRPC chemotherapy regimens (neither HSPC nor CRPC regimens identifiable)	If neither HSPC nor CRPC chemotherapy events were identified in the routine data but a neutropenic event occurred anytime, assume the HES missed the regimen and infer that CRPC chemotherapy was administered.
5	Correctly classify neutropenic event as due to HSPC or CRPC chemotherapy depending on the date of event (<i>when no HSPC or CRPC chemotherapy identified</i>)	If a patient did not have HSPC or CRPC chemotherapy, then infer that the patient did not have sepsis from HSPC or CRPC chemotherapy.
6	Correctly classify neutropenic event as due to HSPC or CRPC chemotherapy depending on the date of event (<i>when one of HSPC or CRPC chemotherapy identified</i>)	If HSPC chemotherapy was administered but not CRPC, and the neutropenic event was identified as occurring ≤25 weeks from randomisation, this was identified as being related to HSPC-chemotherapy. If the patient had HSPC chemotherapy, but the event occurred >25 weeks post-chemotherapy, this neutropenic event was related to CRPC therapy, not HSPC.
7	Correctly classify neutropenic event as due to HSPC or CRPC chemotherapy depending on the date of event (<i>when both HSPC and CRPC regimens were administered</i>)	Identifying if an event was due to HSPC or CRPC chemotherapy if both were administered. The neutropenic event must be >25 weeks after randomisation for a CRPC event. If the event occurred ≤25 weeks after randomisation, then the event was due to HSPC chemotherapy (not CRPC chemotherapy).
8	Identify events not related to chemotherapy and those related	Requirement to identify when the admission event occurred in relation to the chemotherapy. In HES the last data row was identified to calculate a proxy for events being chemotherapy-related. However, due to the additional clinical data in the SACT, enabling a start and end date of regimen to be calculated, a more precise rule was developed assessing the time to the event from the start of the regimen.
9	<i>Inferal rule</i> - Inferring missing HES CRPC chemotherapy regimens (HSPC regimens identifiable)	If no CRPC chemotherapy events were identified but a neutropenic event was detected >25 weeks, then assume that CRPC chemotherapy data were missing and reclassify the event and chemotherapy status as due to CRPC.

Table 21. The nine main algorithm rules created due to the lack of clinical details within the routine data.

Algorithm	Rule								
	1	2	3	4	5	6	7	8	9
HES	+	+	+	-	+	+	+	+	-
HES (Inferred)	+	+	+	+	-	+	+	-	+
HES + SACT	+	+	+	-	+	+	+	+	-

Table 22: Table showing which algorithms utilised which rules.

Many of the rules (table 21) were based upon the below rationale for the timing of chemotherapy regimens and neutropenic events. All algorithm rules were developed to identify events in the absence of clinical variables in both the HES and SACT routine data (for example, hormone-setting and outcomes, including the cause of the neutropenic events).

Chemotherapy timing rules

In the published STAMPEDE docetaxel comparison (12), the median time to starting chemotherapy was 2.4 weeks after randomisation (arm C) and the maximum was not reported, so I estimated this to be 5 weeks post randomisation. The docetaxel regimen should involve six cycles (without dose modification), every three weeks (week = 0-1 #1 docetaxel, week = 3-4 #2 docetaxel, week = 6-7 #3 docetaxel, week = 9-10 #4 docetaxel, week = 12-13 #5 docetaxel, week = 15-16 #6 docetaxel). Therefore, the algorithm utilised the estimate of the maximum time to beginning the STAMPEDE docetaxel regimen (5 weeks) plus the time taken to complete six full cycles (16 weeks), 21 weeks. Therefore, if chemotherapy was identified in the routine data occurring in week 0 – 21 (≤ 21) this was a proxy for HSPC-related chemotherapy and week 22+ (> 21), this was a proxy for CRPC chemotherapy.

When SACT was used in addition to the HES data, SACT was an alternative source of chemotherapy data. Fields were present in SACT to enable identification of start and end dates of treatment (absent in HES). Where the dates of treatment were available, if chemotherapy occurred at randomisation (less than or equal 5 weeks after randomisation), then it was inferred that this was HSPC and not CRPC treatment. This was to confirm that CRPC classified regimens were not just delayed HSPC regimens. If

the chemotherapy began before, or on, week 5 (maximum estimate), the chemotherapy event was classified as being for HSPC.

Neutropenic event timing rules

The distinction of whether the neutropenic event occurred due to HSPC, CRPC chemotherapy or unrelated to chemotherapy was based upon the timing of the neutropenic event compared to the timing of the chemotherapy administration (table 21).

As described in the introduction (section 3.3.2.4), a neutropenic event is most likely to occur within 4 weeks after chemotherapy administration (123). Therefore, with the rationale that most HSPC chemotherapy regimens were completed in 21 weeks, neutropenic events identified less than or equal 25 weeks from randomisation were classified as HSPC-related and not CRPC related. However, if the patient had HSPC chemotherapy, but the event occurred greater than 25 weeks after randomisation, then the neutropenic event was classified as related to CRPC therapy and not HSPC. Varying time intervals for classification of associated neutropenic events were explored (associated event four and twelve-weeks post-chemotherapy).

The last chemotherapy date in the routine data was used to identify if the neutropenic event was unrelated to chemotherapy. The last row in the HES data was used as a proxy for the end of the regimen. However, the length of the regimen was predicted as there was no regimen start or end date available in the HES. If the event occurred greater than 4 weeks after the last chemotherapy, or before it was initiated, then the event was classified as being unrelated to chemotherapy. If the event occurred less than 25 weeks before the last HES-identifiable chemotherapy, or less than or equal 4 weeks after the last HES-identifiable chemotherapy, the event was classified as being related to chemotherapy (≥ -25 or ≤ 4 from the last chemotherapy event).

A SACT alternative was developed because the start and end date of regimens were available. If the event occurred less than 0 weeks from the start of the regimen, this was classified as an unrelated HSPC event, as occurred prior to randomisation into STAMPEDE. If the event occurred greater than or equal to 0 weeks from the first and

less than or equal to 4 weeks after the last cycle, the event was classified as being related to chemotherapy. If no HSPC or CRPC chemotherapy regimens were identified using the HES or SACT algorithms, then any admission events identified were classified as being unrelated to chemotherapy.

Inferal algorithm rules

To identify further regimens, different rules were utilised in addition for the HES algorithm, called the inferal rules. Previously, if no HSPC or CRPC chemotherapy regimens were identified using the HES algorithm, but neutropenic admission events were identified, these events were classified as unrelated to chemotherapy. However, during inferal, even when no chemotherapy regimens were identified, if a neutropenic admission was identified, then this event was classified as a true positive and assigned a HSPC or CRPC setting. This was to account for the missing chemotherapy regimens in HES, leading to misclassification of events.

SACT algorithm rules

Although the SACT algorithm utilised the same rationale for the identification of events as the HES algorithm, rule variations were required in the coding and sequence. Additional data manipulation was also required such as that outlined in the text above and section 3.11.6 (Data preparation).

Analyses	Cohort one (n=44)	Cohort two (n=113)	Cohort three (n=1,573)	Cohort four (n=3,642)
HES (four-week)	✓	✓	✓	✓
HES (inferal) (four-week)	✓	✓	✓	✓
HES (12-week)	✓	✓	✗	✗
HES + SACT	✗	✗	✓	✗

Table 23: *The algorithms developed for the different analyses.*

3.11.7.2 Statistical analyses

I created an algorithm to extract patient characteristics from the STAMPEDE data, including, treatment arm, broad disease grouping, Prostate specific antigen (PSA) at randomisation (median and range), tumour, node and metastases category at

randomisation and age at randomisation (median and range). C. Brawley at the MRC CTU at UCL calculated the age statistics for cohort two, three and four, as I did not have access to date of birth in the STAMPEDE data. This is because it was not necessary compared to the highly identifiable nature of the variable. I calculated the age for cohort one using the UHB QEH clinical noting.

Incidence values were calculated with 95% confidence intervals. Although patients can experience multiple neutropenic events per regimen, only detection of one instance was required per hormone-setting, to enable calculation of incidence at a hormone-sensitivity level (the count of incidence per patient, per regimen, was not examined). This also accounted for duplicate HES reporting.

Odds ratios (OR) with their 95% confidence intervals were calculated to compare the odds of HSPC neutropenic events with those at CRPC. The directionality of affect was as follows; an OR less than one favoured the HSPC chemotherapy group. A forest plot was constructed to compare ORs. The absolute percentage difference for the incidence values were also calculated (HSPC % - CRPC %). Descriptive statistics including sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were calculated where possible, to assess the accuracy of the routine data. However, all were estimates of an unknown truth, with no 100% accurate gold standard (147). Within this chapter, the definitions of the sensitivity, specificity, PPV and NPV are as follows:

- **Sensitivity (se):** The proportion of patients with events who were correctly identified by the algorithm out of the total number of true positives with events (as determined from the clinical noting or the STAMPEDE trial data).
- **Specificity (sp):** The proportion of patients who did not have events who were correctly identified by the algorithm out of the total number of true negatives without events (as determined from the clinical noting or the STAMPEDE trial data).
- **PPV:** The probability of the patients that were identified with the algorithm as having an event (chemotherapy or neutropenic events), truly having an event (as determined from the clinical noting or the STAMPEDE trial data).

- **NPV:** The probability of the patients that were identified with the algorithm as not having an event (chemotherapy or neutropenic events), truly not having an event (as determined from the clinical noting or the STAMPEDE trial data).

Various different processes were undertaken during the cohort one to four analyses (table 24) (figure 15).

Analysis	Cohort 1	Cohort 2	Cohort 3	Cohort 4
Note review comparison	✓	✓*	✗	✗
STAMPEDE trial events comparison	✓	✗	✓	✓
SACT use	✓	✓*	✓	✗
Algorithms run & incidence calculated	✓	✓	✓	✓
Odds ratios, p-values, CIs calculated	✓	✓	✓	✓
12-week event detection	✓	✓	✗	✗
Explicit 'sepsis-only' coding analysis	✗	✗	✓	✓

Table 24: A summary of the analysis completed in each cohort.

* = Targeted note review or SACT comparison to confirm events detected by different algorithms.

To summarise, during all cohort analyses, the algorithms (table 24) were run down the HES data (plus SACT where required) to calculate the incidence values and odds ratios (OR).

In addition, during cohort 1 analyses, routine data coding quality was assessed, an Excel spreadsheet was designed to compare the datasets (what was extracted can be seen in table 15). Note review was undertaken of the clinical noting to identify chemotherapy-related events as occurring whilst, or shortly after, chemotherapy. Cohort 2 analyses included investigation into extra chemotherapy events identified during inferral. These were queried in the clinical noting via targeted note review and were compared to the SACT data.

During cohort 3 analyses the HES neutropenic event coding was analysed for a sample of 70 events (identified using the 'SACT-HES' algorithm) and classified as sepsis-only or neutropenia-only (agranulocytosis) coding (table 16). All STAMPEDE HSPC arm C/E patients (docetaxel) (N=200) were compared to the SACT data and the SACT accuracy was calculated. During the cohort 4 analyses the impact on incidence of restricting

analysis to explicit 'sepsis/septic shock' coding was assessed (table 16). In addition, all of the algorithm detected HSPC neutropenic events were compared to the STAMPEDE documented events (febrile neutropenia events, plus grade 3+ neutropenia events). Including those with and without an explicit admission documented in the STAMPEDE data; accuracy was assessed by year. The ability of the SACT to detect HSPC STAMPEDE chemotherapy data were also assessed (figure 15).

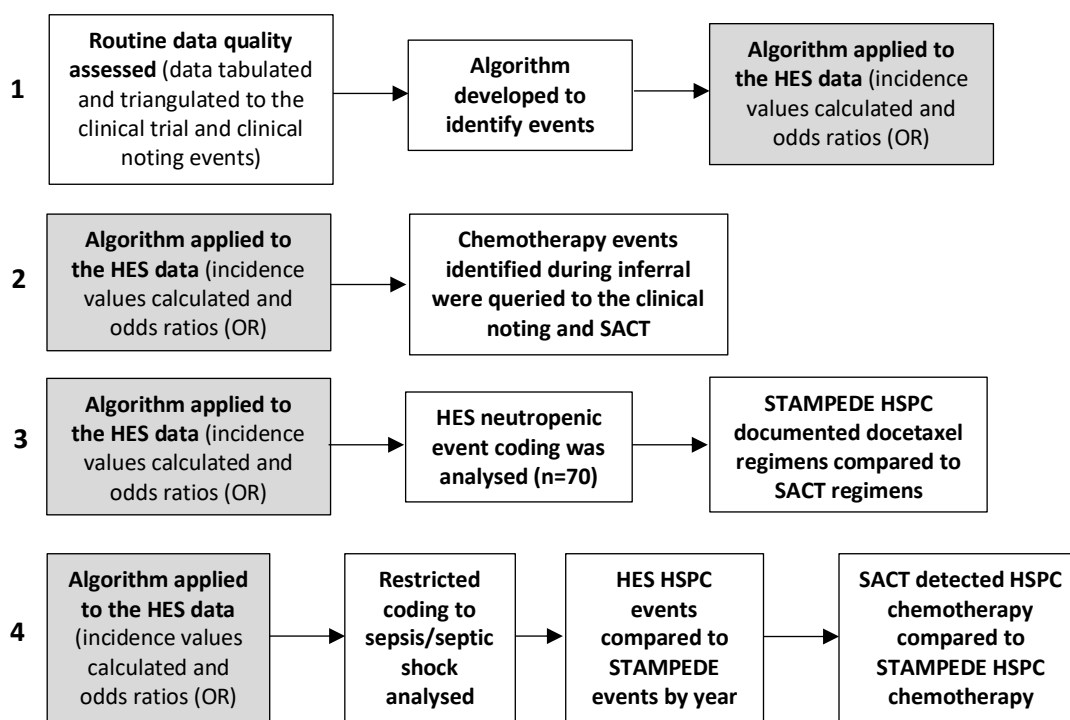


Figure 15: The various processes undertaken in the stage one to four cohort analyses.

3.12 Results

3.12.1 Routine data linkage statistics

8,673 English unique patient IDs were sent to PHE for linkage, 227/8673 (2.6%) failed linkage due to an invalid NHS number. Of the remaining 8,446 patients, 7863/8446 (93.1%) had linked HES data and 4156/8446 (49.2%) had linked SACT data. When restricted to patients randomised after April 2012 (when SACT began collecting), 976/1956 (49.9%) patients had linked SACT data. Those not linked were due to further missing or inaccurate routine data records.

Note: During a consent audit by the MRC CTU at UCL in June 2019, additional patients were identified in the linked dataset who had not given consent for their routine data to be used. As soon as I was made aware of this issue, up to the final analyses cut-off, I excluded these patients from further analyses and removed them from the study cohorts. However, as discussed in section 6.1.1.3 (overall Discussion), currently further investigation is being undertaken which may confirm that these patients remain eligible.

3.12.2 Patient characteristics

The patient characteristics for each analysis cohort are shown in table 25.

Characteristic		Number of patients (% of cohort)			
		Cohort 1 (N=44)	Cohort 2 (N=113)	Cohort 3 (N=1,573)	Cohort 4 (N=3,642)
STAMPEDE Treatment arm	A*	9 (20%)	31 (27%)	636 (40%)	1265 (35%)
	B*	8 (18%)	14 (12%)	95 (6%)	413 (11%)
	C*^	6 (14%)	16 (14%)	97 (6%)	408 (11%)
	D*	6 (14%)	11 (10%)	0 (0%)	206 (6%)
	E*^	9 (20%)	16 (14%)	103 (7%)	424 (12%)
	F*	6 (14%)	13 (12%)	0 (0%)	206 (6%)
	G*	0 (0%)	12 (11%)	642 (41%)	720 (20%)
Broad disease grouping	Newly diagnosed N0M0	10 (23%)	26 (23%)	415 (26%)	843 (23%)
	Newly diagnosed N+M0	3 (7%)	10 (9%)	306 (19%)	585 (16%)
	Newly diagnosed M1	30 (68%)	73 (65%)	780 (50%)	2025 (56%)
	Previously treated M0	0 (0%)	1 (1%)	33 (2%)	88 (2%)
	Previously treated M1	1 (2%)	3 (3%)	39 (2%)	101 (3%)
Age at randomisation	Median (IQR)	63 (57-69)	64 (71-58)	67 (63-72)	66 (62-71)
	Range	48-80	41-81	39-85	39-94
PSA at randomisation	Median (IQR)	109 (57-348)	92 (37-334)	52 (20-154)	59 (21-175)
	Range	10-5000	3-8028	0.2-21460	0.2-21460
T category at randomisation	T0	0 (0%)	0 (0%)	4 (0%)	10 (0%)
	T1	0 (0%)	2 (2%)	12 (1%)	34 (1%)
	T2	2 (5%)	5 (4%)	141 (9%)	330 (9%)
	T3	27 (61%)	67 (59%)	1085 (69%)	2410 (66%)
	T4	7 (16%)	17 (15%)	255 (16%)	612 (17%)
	TX	8 (18%)	22 (19%)	76 (5%)	246 (7%)
N category at randomisation	N0	13 (30%)	42 (37%)	711 (45%)	1613 (44%)
	N+	22 (50%)	57 (50%)	798 (51%)	1838 (50%)
	NX	9 (20%)	14 (12%)	64 (4%)	191 (5%)
Metastases at randomisation	M0	31 (70%)	37 (33%)	819 (52%)	2126 (58%)
	M1	13 (30%)	76 (67%)	754 (48%)	1516 (42%)

Table 25: Patient characteristics by cohort.

T: Tumour status; TX: Tumour cannot be measured; N: Nodal status; NX: Nodal status cannot be measured; N+: Node positive; N0: Node negative; M: Metastatic status; M1: Metastatic; M0: Non-metastatic; *: Potential CRPC chemotherapy candidate; ^: HSPC chemotherapy administered. See figure 14 for the STAMPEDE schema and the introduction 1.5.1 for further details.

3.12.3 Cohort one

The spread of randomisation across arms A to F enabled analysis of a sample of patients that had been administered chemotherapy upfront (arms C and E, 34% of the whole cohort) and those that had not (66% across arms A, B, D and F). The majority of patients were newly diagnosed with M1 disease and the majority of patients were randomised as T3, N+ and M0. All subgroups for metastatic and nodal status were included.

3.12.3.1 Note review analysis

Note review identified a higher neutropenic event incidence if docetaxel or cabazitaxel was given at relapse for CRPC, compared to HSPC (table 26). 15 patients were

administered docetaxel upfront for HSPC in STAMPEDE, of which 2/15 (13.3%) experienced a neutropenic event, both of which were also documented within the STAMPEDE trial data. 24 patients were administered chemotherapy at CRPC using 27 different chemotherapy regimens (three patients had two CRPC chemotherapy regimens): N=1/27 carboplatin and etoposide; 22/27 docetaxel, 4/27 cabazitaxel. 8/24 (33.3%) patients experienced at least one neutropenic event (table 26) (table 27) (OR: 0.40, 95% CI: 0.07 – 2.14, p=0.28) (table 28). For all eight events, the patients had most recently been administered docetaxel or cabazitaxel, not carboplatin and etoposide. There were also multiple neutropenic events per patient (14 in total), 9/14 (64.3%) events appeared to be related to docetaxel administration and 5/14 (35.7%) appeared to be related to Cabazitaxel administration.

Of the 22 patients that were administered docetaxel (other drugs were excluded from analysis) at relapse (CRPC), 6/22 (27.3%, 95% CI: 8.7 – 45.9) appeared to develop a neutropenic event related to the administration (table 26) (OR: 0.49, 95% CI: 0.09 – 2.76, p=0.42) (table 28).

Overall cohort	
Total patients	44
HSPC events	
Number of patients administered a chemotherapy regimen (docetaxel)	15
Number of patients that experienced a related neutropenic admission	2
Incidence - patients on docetaxel developing a neutropenic event (%)	13.3
CRPC events (all chemotherapy)	
Number of patients administered a chemotherapy regimen (not restricted to docetaxel)	24
Number of patients that experienced a related neutropenic admission	8
Incidence - patients on chemotherapy developing a neutropenic event (%)	33.3
Neutropenic events due to docetaxel	9/14 (64%)
Neutropenic events due to cabazitaxel	5/14 (36%)
CRPC events (docetaxel only)	
Number of patients administered docetaxel at relapse	22
Number of patients that experienced a related neutropenic admission	6
Incidence - patients on docetaxel developing a neutropenic event (%)	27.3

Table 26: The number of patients on HSPC and CRPC chemotherapy regimens and the incidence of neutropenic events.

3.12.3.2 Routine data analysis

3.12.3.2.1 Chemotherapy data quality

The HES data identified 0/15 (sensitivity: 0.00) HSPC chemotherapy regimens. Therefore, the HES-only algorithm was developed by merging the HES with the STAMPEDE trial HSPC regimens. Hence, by definition, the algorithm would identify 100% of HSPC regimens correctly, as such, 15/15 (sensitivity: 1.00) were then identified. In the CRPC setting, 21/24 patients who were administered at least one regimen (all chemotherapy drugs) were identified in the HES (identification required at least one cycle to be present in these data) (figure 16) (the accuracy of chemotherapy coding was further assessed in chapter 5). All patients identified by the algorithm as having been administered CRPC chemotherapy, were true positive events, compared to the reference (PPV: 1.00). The three missing HES regimens were identified during the inferral analysis because a neutropenic admission event was present (see the inferral algorithm rules, section 3.11.7.1) (sensitivity: 1.00, specificity: 1.00, PPV: 1.00, NPV: 1.00).

	(Reference) CRPC chemotherapy (patients)	(Reference) No CRPC chemotherapy (patients)	
(Algorithm) CRPC chemotherapy (patients)	21	0	PPV 1.00
(Algorithm) No CRPC chemotherapy (patients)	3	20	NPV 0.87
	Se 0.88	Sp 1.00	

Figure 16: Number of CRPC chemotherapy regimens identified by the HES-only algorithm, prior to inferring missing regimens.

Of the 27 CRPC regimens, only 5/27 were prescribed after April 2012 (when SACT chemotherapy data began collection). SACT was able to identify 4/5 (80%) which were all prescribed after 2014.

3.12.3.2.2 Neutropenic event data quality

At a patient level, 7/10 (sensitivity: 0.70) patients were correctly identified as experiencing a neutropenic event by the algorithm (for both HSPC and CRPC regimens) and 33/34 patients were correctly identified as not experiencing a neutropenic event (specificity: 0.97) (PPV: 0.88, NPV: 0.92) (figure 17).

	(Reference) Sepsis (patients)	(Reference) No sepsis (patients)	
(Algorithm) Neutropenic events (patients)	7	1	PPV 0.88
(Algorithm) No neutropenic events (patients)	3	33	NPV 0.92
	Se 0.70	Sp 0.97	

Figure 17: The number of neutropenic events identified by the algorithms, prior to inferring missing events.

After inferring missing chemotherapy, 9/10 (sensitivity: 0.90) patients with a neutropenic event were identified, and 32/34 patients were correctly identified by the algorithm as not having a neutropenic event (specificity: 0.94) (PPV: 0.82, NPV: 0.97). All events detected were correctly allocated to HSPC or CRPC chemotherapy (table 27).

Criteria	Reference	HES (%)	HES (infernal) (%)
HSPC events			
Chemotherapy detection			
Number of patients identified as being administered HSPC docetaxel	15	15/15 (100)	15/15 (100)
Neutropenic event detection			
Number of patients correctly identified as experiencing a HSPC docetaxel-related neutropenic event	2	2/2 (100)	2/2 (100)
Number of false negative (missed) neutropenic events (event and patient level)	0	0	0
Number of false positive detected neutropenic events (event and patient level)	0	0	0
Number of patients correctly identified as not experiencing a HSPC docetaxel-related neutropenic event	13	13/13 (100)	13/13 (100)
<i>Incidence of HSPC docetaxel neutropenic events (%)</i>	13.3	13.3	13.3
CRPC events			
Chemotherapy detection			
Number of patients identified as being administered CRPC chemotherapy	24	21/24 (88)	24/24 (100)
Number of patients with missed CRPC chemotherapy	0	3/24 (13)	0/24 (0)
Number of patients correctly identified as having CRPC chemotherapy	24	21/21 (100)	24/24 (100)
False positive detected CRPC chemotherapy regimens	0	0	0
Neutropenic event detection			
Number of patients identified as experiencing a chemotherapy neutropenic event	8	6	9
Number of patients correctly identified as experiencing a CRPC chemotherapy-related neutropenic event	8	5/8 (63)	7/8 (88)
Number of patients with false negative (missed) neutropenic events	0	3/8 (38)	1/8 (13)
Number of patients correctly identified as not experiencing a CRPC chemotherapy-related neutropenic event	36	35/36 (97)	34/36 (94)
Number of patients with false positive CRPC neutropenic events	0	1	2
Number of correctly identified neutropenic events (event level)	14	6/14 (43)	8/14 (57)
Number of false negative (missed) neutropenic events (event level)	0	8/14 (57)	6/14 (43)
<i>Incidence of CRPC chemotherapy neutropenic events (%)</i>	33.3	28.6	37.5
HSPC and CRPC events			
Number of neutropenic events identified correctly	16	8/16 (50)	10/16 (63)
Number of false negative (missed) neutropenic events	0	8/16 (50)	6/16 (38)
Number of patients with correctly identified neutropenic events	10	7/10 (70)	9/10 (90)

Table 27: The number of events and patients identified using the routine data algorithm compared to the reference note review and the impact of inferring missing chemotherapy regimens. Percentages are rounded to one decimal place. Grey: highlighting the reference values.

Overall, utilising the 'HES algorithm', a 13.3% (95% CI: 0.00 – 30.5) neutropenic event incidence was identified, appearing to be related to HSPC docetaxel and a 28.6% (95% CI: 9.2 – 47.9) neutropenic event incidence upon CRPC chemotherapy (OR: 0.47, 95% CI: 0.08 – 2.64, $p=0.39$). If missing chemotherapy events were inferred, the CRPC incidence increased to 37.5% (95%, CI: 18.1 – 56.9) (table 27) (OR: 0.36, 95% CI: 0.07 – 1.87, $p=0.21$) (table 28). The HSPC neutropenic event incidence remained unchanged due to identifying all HSPC regimens through linkage to the STAMPEDE database. The ORs for experiencing a neutropenic event due to HSPC compared to CRPC chemotherapy can be seen in table 28. Despite the OR favouring an increased neutropenic event incidence for CRPC, the ORs were not significant ($p>0.05$), suggesting no differences were present between the groups. Due to the small cohort, the incidence estimate CIs also overlapped (table 28).

	HSPC docetaxel neutropenic event incidence (%) (95% CI)	CRPC chemotherapy neutropenic event incidence (%) (95% CI)	OR (95% CI)	OR p- value
Note review (all chemotherapy)	13.3 (0.00 - 30.5)	33.3 (14.5 – 52.2)	0.40 (0.07–2.14)	0.28
Note review (docetaxel only)	13.3 (0.00 - 30.5)	27.3 (8.7 – 45.9)	0.49 (0.09 – 2.76)	0.42
HES algorithm	13.3 (0.00 - 30.5)	28.6 (9.2 – 47.9)	0.47 (0.08 – 2.64)	0.39
HES algorithm (inferred)	13.3 (0.00 - 30.5)	37.5 (18.1 – 56.9)	0.36 (0.07 – 1.87)	0.21

Table 28: Cohort one analysis HSPC and CRPC rates and the corresponding ORs.

When widening the classification window for developing a chemotherapy-related neutropenic event from within or equal to, four weeks after chemotherapy, to twelve, one additional event was classified. However, this patient had already been classified as experiencing a CRPC event, so did not impact the patient level incidence (see table 29).

	HES algorithm (≤4 weeks)	HES algorithm (≤12 weeks)	HES algorithm (inferal) (≤4 weeks)
Number of patients administered HSPC docetaxel	15	15	15
Number of patients experiencing a neutropenic event	2/15	2/15	2/15
Neutropenic event incidence (HSPC) (%)	13.3	13.3	13.3
Number of patients administered CRPC chemotherapy	21	21	24
Number of patients experiencing a neutropenic event	6/21	6/21	9/24
Neutropenic event incidence (CRPC) (%)	28.6	28.6	37.5

Table 29: The incidence of neutropenic admissions from HSPC or CRPC chemotherapy, identified using the HES algorithms (\pm inferal) with the neutropenic event inclusion of ≤ 4 or ≤ 12 weeks after the last chemotherapy event.

	True positive events	False positive events	False negative events	True negative events	Sensitivity	PPV	Specificity	NPV	Incidence of HSPC neutropenic events (%)	Incidence of CRPC neutropenic events (%)		
Reference												
Neutropenic events (CRPC) (patient level)	8	0	0	36	-	-	-	-	13.3	33.3		
Neutropenic events (HSPC) (patient level)	2	0	0	42	-	-	-	-				
Total neutropenic events (HSPC, CRPC) (patient level)	10	0	0	34	-	-	-	-				
Neutropenic events (CRPC) (event level)	14	0	0	NA	-	-	-	-				
Neutropenic events (HSPC) (event level)	2	0	0	NA	-	-	-	-				
Number of CRPC chemotherapy regimens (patient level)	24	0	0	20	-	-	-	-	13.3	28.6		
Number of HSPC chemotherapy regimens (patient level)	15	0	0	29	-	-	-	-				
Routine data algorithm (HES)												
Neutropenic events (CRPC) (patient level)	5	1	3	35	0.63	0.83	0.97	0.92				
Neutropenic events (HSPC) (patient level)	2	0	0	42	1.00	1.00	1.00	1.00				
Total neutropenic events (HSPC, CRPC) (patient level)	7	1	3	33	0.70	0.88	0.97	0.92	13.3	28.6		
Neutropenic events (CRPC) (event level)	6	1	8	NA	0.43	0.86	NA	NA				
Neutropenic events (HSPC) (event level)	2	1	0	NA	1.00	0.67	NA	NA				
Number of CRPC chemotherapy regimens (patient level)	21	0	3	20	0.88	1.00	1.00	0.87				
Number of HSPC chemotherapy regimens (patient level)	15	0	0	29	1.00	1.00	1.00	1.00				
Infernal (HES)												
Neutropenic events (CRPC) (patient level)	7	2	1	34	0.88	0.78	0.94	0.97	13.3	37.5		
Neutropenic events (HSPC) (patient level)	2	0	0	42	1.00	1.00	1.00	1.00				
Total neutropenic events (HSPC, CRPC) (patient level)	9	2	1	32	0.90	0.82	0.94	0.97				
Neutropenic events (CRPC) (event level)	8	3	6	NA	0.57	0.73	NA	NA				
Neutropenic events (HSPC) (event level)	2	1	0	NA	1.00	0.67	NA	NA				
Number of CRPC chemotherapy regimens (patient level)	24	0	0	20	1.00	1.00	1.00	1.00	13.3	37.5		
Number of HSPC chemotherapy regimens (patient level)	15	0	0	29	1.00	1.00	1.00	1.00				

Table 30: Detection of neutropenic events.

‘-’ = the sensitivity, specificity, PPV and NPV were not documented for the reference, as would be 1.00 for all, due to being the reference; NA = at an individual event level there was an infinite number of times when neutropenic events did not occur, hence calculations of specificity and NPV could not be undertaken.

3.12.4 Cohort two

The greatest number of patients were randomised to arm A (27.4%), with a relatively equal distribution across the other arms B – G (lowest D, 9.7%; highest C/E, 14.2%) (table 25).

Using the routine data algorithms, a higher incidence of neutropenic events was identified for chemotherapy at CRPC than HSPC, although, non-significant ORs were calculated. 4/32 (12.5%, 95% CI: 1.0 – 24.0) of the patients that had HSPC chemotherapy developed a neutropenic event and 9/47 (19.1%, 95% CI: 7.9 – 30.4) with CRPC chemotherapy (OR; 0.65, 95% CI: 0.19 – 2.30, p=0.51). When widening the detection window from four to twelve weeks, one more event was identified, increasing the CRPC neutropenic event incidence to 10/47 (21.3%, 95% CI: 9.6 – 33.0) (table 31).

	HES (≤4 weeks)	HES (≤12 weeks)	HES (≤4 weeks) (inferal)	HES (validated chemotherapy events)	HES (+ SACT data)
Number of patients (HSPC Docetaxel)	32	32	32	32	32
Number of HSPC docetaxel neutropenic events (patient level)	4	4	4	4	4
Number of patients (CRPC chemotherapy)	47	47	52	53	50
Number of CRPC chemotherapy neutropenic events (patient level)	9	10	18	15	12
HSPC docetaxel neutropenic event incidence (%) (95% CI)	12.5 (1.0 – 24.0)	12.5 (1.0 – 24.0)	12.5 (1.0 – 24.0)	12.5 (1.0 – 24.0)	12.5 (1.0 – 24.0)
CRPC chemotherapy neutropenic event incidence (%) (95% CI)	19.1 (7.9 – 30.4)	21.3 (9.6 – 33.0)	34.6 (21.7 – 47.5)	28.3 (16.2 – 40.4)	24.0 (12.2 – 35.8)

Table 31: The incidence of neutropenic events by the timing of chemotherapy (HSPC, CRPC) identified using the routine data algorithms (varying time intervals for event association, clinical noting validation of missing HES regimens and the SACT analysis), during cohort two analysis.

Algorithm	OR (95% CI)	OR p-value	Interpretation
HES (≤ 4 weeks)	0.65 (0.19 – 2.30)	0.51	The odds of a neutropenic event with HSPC chemotherapy was 0.65 of the odds at CRPC (non-significant) The CIs were wide. Crossed the null hypothesis that the odds ratio was equal to 1, with no difference between groups
HES (inferred)	0.36 (0.11 – 1.16)	0.08	The odds of a neutropenic event with HSPC chemotherapy was 0.36 of the odds at CRPC (non-significant) The CIs were wide. Crossed the null hypothesis that the odds ratio was equal to 1, with no difference between groups
HES (validated chemotherapy events)	0.44 (0.13 – 1.45)	0.17	The odds of a neutropenic event with HSPC chemotherapy was 0.44 of the odds at CRPC (non-significant) The CIs were wide. Crossed the null hypothesis that the odds ratio was equal to 1, with no difference between groups

Table 32: Odds ratios comparing the incidence of experiencing a neutropenic event by chemotherapy setting.

ORs <1 favour the HSPC group.

A further 9 events were identified during inferred of missing routine data chemotherapy events, increasing the CRPC chemotherapy neutropenic event rate to 18/52 (34.6%, 95% CI: 21.7 – 47.5) (table 31) (OR; 0.36, 95% CI: 0.11 – 1.16, p=0.08) (table 32).

Note review determined that 3/9 of the extra neutropenic events identified by inferred, were missed previously due to the fact that they occurred after the detection window of four or twelve weeks from the last identifiable chemotherapy event. The remaining 6/9 were missed due to missing chemotherapy HES data (appendix, table 78). Note review analysis confirmed that chemotherapy was administered in all six of the cases but was not coded in the HES (table 33). Therefore, the true rate of neutropenic events at CRPC was actually closer to the inferred value, increasing the incidence to 15/53 (28.3%, 95% CI: 16.2 – 40.4) (table 31) (OR: 0.44, 95% CI: 0.13 – 1.45, p=0.17) (table 32).

ID	Routine data detected neutropenic event (weeks post STAMPEDE randomisation)	Chemotherapy drug identified in the clinical noting	Event chemotherapy related?	Further detail
1	188	Docetaxel	✓	TRAPEZE trial - not in HES. Event occurred during the chemotherapy regimen.
9	358	Docetaxel	✓	Two events identified. Whilst on docetaxel; event occurred during the chemotherapy regimen.
9	358	Cabazitaxel	-	Two events identified. Not whilst on cabazitaxel; event did not occur during the following cabazitaxel regimen, hence, potential duplicate
10	82	Docetaxel	✓	TRAPEZE trial - not in HES. Event occurred during the chemotherapy regimen.
12	40	Docetaxel	✓	TRAPEZE trial - not in HES. Event occurred during the chemotherapy regimen.
17	101	Cabazitaxel	✓	Cabazitaxel event. Event occurred during the chemotherapy regimen.
18	191	Docetaxel	✓	Docetaxel event. Event occurred during the chemotherapy regimen.

Table 33: Assessment of the six extra events that were identified during the inferal analysis. Further details are in the appendix table 78 with corresponding IDs; - = potential duplicate.

Of these, 3/6 (50%) were identifiable in the SACT data but the other 3/6 (50%) occurred prior to SACT data collection, so could not be assessed. These three extra events identified using SACT, occurred in 2015, 2016 and 2017. Hence, the addition of SACT enabled detection of 12/50 (24.0%, 95% CI: 12.2 – 35.8) neutropenic events at CRPC (table 31).

3.12.5 Cohort three

The majority of patients were randomised into arms A (40%) and G (41%), with fewer on B (6%), C (6%) and E (7%). No patients were analysed in D or F, due to the SACT data collection (see methods, figure 12). The broad disease grouping showed that the majority of patients were newly diagnosed metastatic (M1) patients, with a low percentage of those previously treated. Most patients were diagnosed with T3, N+, and M0 disease at randomisation (table 25).

Using the 'SACT-HES' algorithm enabled analysis restricted to docetaxel. A similar incidence of events at HSPC and CRPC were identified. 41/200 (20.5%, 95% CI: 14.9 – 26.1) developed a neutropenic event with HSPC chemotherapy and 60/297 (20.2%, 95% CI: 15.6 – 24.8) with CRPC chemotherapy (OR: 1.01, 95% CI: 0.66 – 1.57, $p=0.95$) (table 34). Although the OR showed a difference between the groups, the p -value was non-significant. The HES algorithm identified 41/200 (20.5%, 95% CI: 14.9 – 26.1) patients with a HSPC chemotherapy neutropenic event and 51/378 (13.5%, 95% CI: 10.9 – 16.9) with a CRPC neutropenic event. However, the ORs were again non-significant (OR: 1.52, 95% CI: 0.97 – 2.37, $p=0.06$) (table 34). With inferral, a statistically significant OR was identified, 163/448 events occurred in the CRPC group (36.4%, 95% CI: 31.9 – 40.8) (OR: 0.56, 95% CI: 0.38 – 0.82, $p=0.03$), compared to 20.5% in the HSPC group (table 34).

81 extra chemotherapy regimens were coded for in HES than in SACT (HES is not restricted to docetaxel only). In addition, using the HES inferral algorithm, 70 further chemotherapy regimens were identified, than using the non-inferral algorithm. 151 more events were identified using the HES inferral algorithm than the SACT algorithm.

Algorithm	OR (95% CI)	p-value	Interpretation
HES (+SACT)	1.01 (0.66 – 1.57)	0.95	The odds at HSPC were similar to CRPC (non-significant) The CIs were wide. Crossed the null hypothesis that the odds ratio was equal to 1, with no difference between groups
HES	1.52 (0.97 – 2.37)	0.06	The odds in the HSPC group were 1.52 of the odds at CRPC (non-significant) The CIs were wide. Crossed the null hypothesis that the odds ratio was equal to 1, with no difference between groups
HES (inferral)	0.56 (0.38 – 0.82)	0.003	The odds in the HSPC group were 0.56 of the odds at CRPC (significant result) The CIs were narrower and did not cross 1, suggesting 95% confidence that the odds were lower in the HSPC setting

Table 34: ORs comparing the incidence of experiencing a neutropenic event by chemotherapy setting.

In a sample of 70 patients that experienced a neutropenic admission, only 7/18 of the eligible codes (table 16) were identified in the routine data. Of these, only 8/70 (11.4%) events were coded as agranulocytosis alone, with no explicit ‘sepsis’ coding. The majority of events 62/70 (88.6%) were explicitly coded as ‘sepsis/septic-shock’ related.

SACT data quality was also assessed; SACT detected 83/200 (41.5%) STAMPEDE upfront regimens. The quality did not improve between 2012 and 2013 (table 35).

Year	Events	Detected	%
2012	152	63	41.5
2013	48	20	41.7
Total	200	83	41.5

Table 35: The ability of the SACT data to detect HSPC STAMPEDE docetaxel regimens.

3.12.6 Cohort four

Patients were randomised into all arms A-G, with the majority in arm A (35%), followed by arm G (20%). The majority were newly diagnosed with most patients randomised as T3, N+, and M0 (table 25).

The HES algorithm identified 134/832 (16.1%, 95% CI: 13.6 – 18.6) patients with a HSPC docetaxel neutropenic event and 148/1182 (12.5%, 95% CI: 10.6 – 14.4) with a CRPC chemotherapy neutropenic event (OR: 1.29, 95% CI: 1.00 – 1.65, $p=0.05$) (table 36); however, again the OR was non-significant. Upon inferral, 489/1350 CRPC chemotherapy neutropenic events were identified (36.2%, 95% CI: 33.7 – 38.8) (OR: 0.44, 95% CI: 0.36 – 0.55, $p=1.4e^{-14}$) (table 36). Hence, this inferral suggested there were lower odds of developing a HSPC neutropenic event.

When removing non-explicit sepsis coding from the analysis, 182 unique patients experienced sepsis due to HSPC or CRPC chemotherapy (183 events in total, one patient experienced an event in both settings). The incidence of sepsis-only events using the HES-only algorithm was 69/832 (8.3%, 95% CI: 6.4 – 10.2) at HSPC and 114/1182 (9.6%, 95% CI: 8.0 – 11.3) at CRPC (OR: 0.86, 95% CI: 0.63 – 1.17, $p=0.34$) (table 36). Despite the lower HSPC incidence of events, a non-significant OR was present.

Algorithm	OR (95% CI)	p-value	Interpretation
HES only	1.29 (1.00 - 1.65)	0.05	The odds of a neutropenic event in the HSPC group were 1.29 of the odds in the CRPC group (significant) The confidence interval touched 1, meaning the odds could have been equal in both groups
HES only (inferred)	0.44 (0.36 – 0.55)	1.4e ⁻¹⁴	The odds of a neutropenic event in the HSPC group were 0.44 of the odds in the CRPC group (significant) The confidence intervals were narrower and did not cross 1, suggesting 95% confidence that the odds were lower in the upfront group
HES only (explicit sepsis coding only)	0.86 (0.63 – 1.17)	0.34	The odds of a neutropenic event in the HSPC group were 0.86 of the odds in the CRPC group (non-significant) The confidence intervals were wide. Crossed the null hypothesis that the odds ratio was equal to 1, with no difference between groups

Table 36: ORs comparing the incidence of experiencing a neutropenic event depending on chemotherapy setting.

In this cohort, 83 neutropenic events were documented in the STAMPEDE database as occurring due to HSPC docetaxel, showing an incidence of 83/832 (10.0%, 95% CI; 7.9 – 12.0). These events were documented as: neutropenia, neutropenic sepsis, febrile neutropenia, neutropenia (infection), pyrexia, neutropenic infection, pneumonia and infection. 59/83 (sensitivity: 0.71) STAMPEDE events were identified in the HES with 75 false positives (PPV: 0.44). 749/832 with HSPC chemotherapy did not experience a neutropenic event, HES identified 674/749 without an event (specificity: 0.90) (figure 18, A). When STAMPEDE admission-only events were analysed (52/83 were related to admissions) (figure 18, B), 48/52 (sensitivity 0.92) events were identified. However, the PPV reduced to 0.36, as 86 false positives were identified. HES also identified 11 admission events that were documented by the trial but were not documented as admissions.

A					
	STAMPEDE event	No STAMPEDE event			
HES algorithm +ve	59	75	134	PPV	44.0
HES algorithm -ve	24	674	698	NPV	96.6
	83	749	832		
	Se	Sp			
	71.1	90.0			
B					
	STAMPEDE event	No STAMPEDE event			
HES algorithm +ve	48	86	134	PPV	35.8
HES algorithm -ve	4	694	698	NPV	99.4
	52	780	832		
	Se	Sp			
	92.3	89.0			

Figure 18: The sensitivity, specificity, PPV and NPV for detecting neutropenic events in the routine data compared to the STAMPEDE trial.

A) All STAMPEDE documented neutropenic events grade 3+ or any febrile neutropenia grade 1+;

B) Events identified in A that led to an admission identifiable in the STAMPEDE database.

The accuracy of HES neutropenic event detection compared to the STAMPEDE documented events varied over time. The accuracy decreased to 30% in 2009 from 100% in 2006. After 2009 the accuracy increased to 100% (2013) (table 37) (figure 19).

Year	Events	Detected	%
2006	1	1	100.00
2007	6	4	66.67
2008	5	4	80.00
2009	10	3	30.00
2010	14	10	71.43
2011	20	14	70.00
2012	22	18	81.82
2013	5	5	100.00
Total	83	59	71.08

Table 37: The number of events identified in the routine data that were documented by the STAMPEDE trial (all instances of febrile neutropenia, plus grade 3+ neutropenic events).

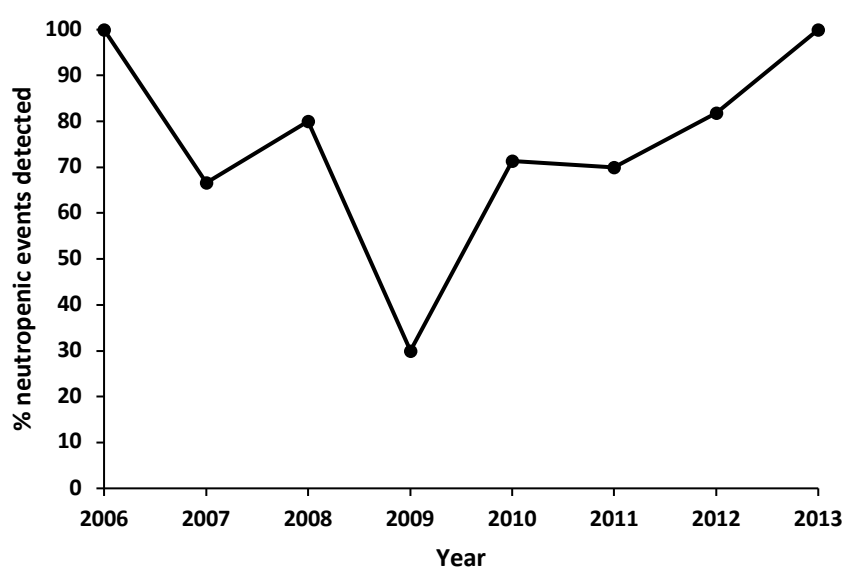


Figure 19: The accuracy of detecting STAMPEDE documented events in the HES data (2006 – 2013).

3.12.7 Overall results

Overall, these analyses do not support that HSPC docetaxel has a higher severe neutropenic event risk than CRPC use. Rates varied by the algorithm used, but these data suggest CRPC docetaxel leads to a similar neutropenic event rate than use in HSPC. However, caveats to this study can be seen below. The incidence values for the largest cohorts can be seen in table 38 and are illustrated in a forest plot in figure 20. Table 39 summarises all stages.

Algorithm	Cohort number	Regimen number HSPC, CRPC	Number HSPC neutropenic events (%)	Number CRPC neutropenic events (%)	Absolute % difference HSPC-CRPC	OR (95% CI)	p-value
HES + SACT	1573	200, 297	41 (21)	60 (20)	+1	1.01 (0.66–1.57)	0.95
HES-only (all events)	3642	832, 1182	134 (16)	148 (13)	+3	1.29 (1.00–1.65)	0.05
HES-only (inferral)	3642	832, 1350	134 (16)	489 (36)	-20	0.44 (0.36–0.55)	<0.001
HES-only (sepsis-only)	3642	832, 1182	69 (8)	114 (10)	-2	0.86 (0.63–1.17)	0.34

Table 38: Incidence values for the largest cohort for each algorithm.

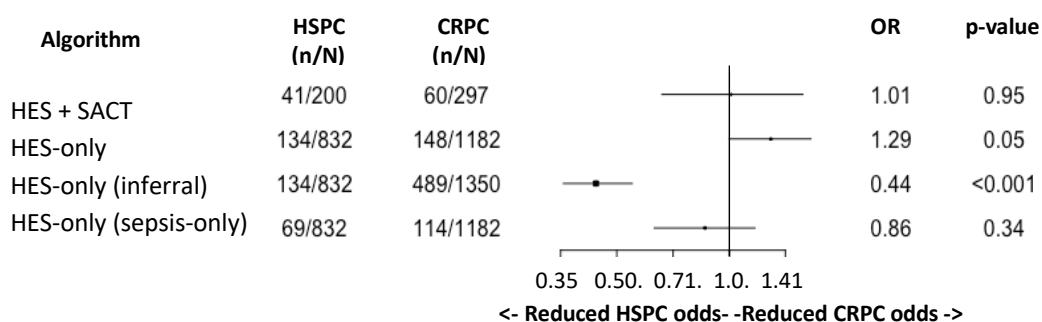


Figure 20: Forest plot showing the number of events (n) by the number of patients (N) in each analysis. ORs are displayed, an OR < 1 favours the HSPC group.

Although a different measure of neutropenic events was developed compared to standard trial definitions and measures, this analysis enabled a direct comparison of major neutropenic events occurring in the different settings. The routine data had limitations such as missingness and lack of clinical details impacting event detection (for example, in the CRPC group in the HES only analyses). However, in the absence of standard randomised follow-up data, routine data were able to test the hypothesis. However, due to these limitations, this study was not able to generate conclusive evidence and further investigation is required.

Cohort	Algorithm	Number of HSPC regimens, number of HSPC neutropenic events	Number of CRPC regimens, number of CRPC neutropenic events	% HSPC chemotherapy event incidence (95% CI)	% CRPC chemotherapy event incidence (95% CI)	% difference HSPC – CRPC incidence	OR (95% CI)	OR p-value
1	Note review (doca + caba)	15, 2	24, 8	13.3 (-3.9 - 30.5)	33.3 (14.5 – 52.2)	-20.0	0.40 (0.07 – 2.14)	0.28
	Note review (doca)	15, 2	22, 6	13.3 (-3.9 - 30.5)	27.3 (8.7 – 45.9)	-14.0	0.49 (0.09 – 2.76)	0.42
	HES	15, 2	21, 6	13.3 (-3.9 - 30.5)	28.6 (9.2 – 47.9)	-15.3	0.47 (0.08 – 2.64)	0.39
	HES (inferral)	15, 2	24, 9	13.3 (-3.9 - 30.5)	37.5 (18.1 – 56.9)	-24.2	0.36 (0.07 – 1.87)	0.21
	HES	32, 4	47, 9	12.5 (1.0 – 24.0)	19.1 (7.9 – 30.4)	-6.6	0.65 (0.19 – 2.30)	0.51
2	HES (inferral)	32, 4	52, 18	12.5 (1.0 – 24.0)	34.6 (21.7 – 47.5)	-22.1	0.36 (0.11 – 1.16)	0.08
	HES (+ validated inferred chemotherapy)	32, 4	53, 15	12.5 (1.0 – 24.0)	28.3 (16.2 – 40.4)	-15.8	0.44 (0.13 – 1.45)	0.17
	HES (+ SACT)	200, 41	297, 60	20.5 (14.9 – 26.1)	20.2 (15.6 – 24.8)	0.3	1.01 (0.66 – 1.57)	0.95
3	HES	200, 41	378, 51	20.5 (14.9 – 26.1)	13.5 (10.9 – 16.9)	7.0	1.52 (0.97 – 2.37)	0.06
	HES (inferral)	200, 41	448, 163	20.5 (14.9 – 26.1)	36.4 (31.9 – 40.8)	-15.9	0.56 (0.38 – 0.82)	0.003
	HES	832, 134	1182, 148	16.1 (13.6 – 18.6)	12.5 (10.6 – 14.4)	3.6	1.29 (1.00 – 1.65)	0.05
4	HES (inferral)	832, 134	1350, 489	16.1 (13.6 – 18.6)	36.2 (33.7 – 38.8)	-20.1	0.44 (0.36 – 0.55)	<0.001
	HES (sepsis only)	832, 69	1182, 114	8.3 (6.4 – 10.2)	9.6 (8.0 – 11.3)	-1.3	0.86 (0.63 – 1.17)	0.34
	STAMPEDE	832, 83	NA	10.0 (7.9 – 12.0)	NA	NA	NA	NA

Table 39: The number of events identified by the algorithm and dataset analysed, including the percentage incidence by timing of chemotherapy administration. Black = the largest algorithm cohorts, being: SACT = cohort three; HES (with and without inferral) = cohort four; note review = cohort one.

3.13 Discussion

3.13.1.1 Main results summary

During cohort one analyses (section 3.12.3), the note review and the routine data algorithms found a higher incidence of chemotherapy neutropenic admission events at CRPC. However, overlapping confidence intervals and non-significant odds ratios suggested no differences may have been present between the two groups (table 28). Note review confirmed that missing HES events (chemotherapy and neutropenic admissions) were impacting the incidence as the reference value lay between the HES algorithm with and without inferring missing chemotherapy regimens (table 28). This showed that inferring events enabled detection of more true positive events, but did also lead to more false positives, which after investigation, were identified as neutropenic events unrelated to chemotherapy.

Due to missing HSPC HES chemotherapy data, the algorithm required linkage to the STAMPEDE trial arm data, to ensure correct identification of the HSPC incidence. However, due to the lack of CRPC data in the trial database, hence the requirement to use routine data for the analyses, it was not possible to identify the CRPC chemotherapy drug administered. Because the drug name could not be identified in the HES data, the CRPC incidence values included the reporting of all chemotherapy drugs. However, SACT data, although limited by timeliness (restricting the analyses), was able to enhance chemotherapy event detection and could identify drug name. Despite this, it was thought reasonable to infer that unspecified CRPC chemotherapy regimens were docetaxel as there was usually only one chemotherapy drug available during this time. Chemotherapy could have been administered for another cancer, but this is estimated to have had little impact on the CRPC estimates. Often patients with other *previous or current cancers* are excluded from the STAMPEDE trial (72), hence the greatest concern is confounding chemotherapy for a second primary after the diagnosis of prostate cancer (chemotherapy for metastatic prostate cancer is certainly an event of interest).

In the larger single-site analyses (cohort two, section 3.12.4), again the HES algorithms identified a higher incidence of neutropenic events with CRPC chemotherapy (table 31).

The HSPC and CRPC incidence values were converging inversely to sample size, compared to the estimates in cohort one. The confidence intervals still overlapped and the ORs were non-significant (table 32). Chemotherapy data were again missing from HES and due to this, the value in the CRPC setting was confirmed (by note review) to be closer to the inferred value (table 31). SACT was able to enhance the detection of chemotherapy events including specifying drug name, although this was only possible for events after April-2012 but was not mandatory to collect until July-2014, leading to sporadic reporting.

During the cohort three analyses (section 3.12.5), the HES plus SACT algorithm identified similar odds at HSPC and CRPC, although a non-significant odds ratio was present. Although SACT data were seen to enhance detection of CRPC events, early SACT was not able to identify HSPC events with high accuracy. When using the HES-only algorithm, a marginally lower incidence of neutropenic events was seen in the CRPC setting, compared to HSPC, but the inferred incidence of events was higher at CRPC.

Most HES events were classified as 'sepsis/septic shock related', but some were coded as neutropenia-only events at admission. Thus, less severe events may have led to an admission, impacting rates.

When utilising the HES algorithm in the cohort four analyses (section 3.12.6), the HSPC and CRPC incidence rates further converged and the incidence of neutropenic events across groups were similar but marginally increased at HSPC. However, when restricting the analysis to 'sepsis-only' coding, the incidence became lower in the HSPC setting, hence the odds ratio favouring the HSPC group (table 36). As identified during the previous cohort analyses, the HES data were missing chemotherapy events, and when inferred, the incidence at CRPC was higher than at HSPC. It is suspected that the true value lies between both indicators.

In addition, the sensitivity of HES to detect STAMPEDE neutropenic event admissions was high, but the PPV was low, due to potential false positive events (figure 18). Although there were many routine data false positive HSPC admission events, STAMPEDE may have missed events and thus the routine data may have been

identifying missed trial events (being reclassified as true positives). I have proposed that the MRC CTU at UCL send queries to individual sites to confirm these events.

Due to the limitations of the routine data proposed here, the results presented are hypothesis generating and thus not conclusive; hence, further investigation is required.

3.13.2 Hypothesis generation

Note review was undertaken to create a reference for routine data feasibility assessment and enabled the calculation of the incidence of neutropenic admissions in the HSPC and CRPC settings. This note review took a long time to ensure accuracy and completeness of events and hence, was highly resource intensive. Due to the HES data lacking clinical details (e.g. drug name), during the note review the incidence was calculated for both docetaxel-alone and for all chemotherapy drugs, to enable comparison to the HES-derived outcomes. When undertaking the note review, it was identified that if chemotherapy was administered in the CRPC setting, the incidence of neutropenic admissions were higher than if administered at HSPC diagnosis (table 26). When restricting to docetaxel only, the CRPC incidence was marginally lower than that for all chemotherapy drugs; however, still remained much higher than in the HSPC group (table 26). This suggested that other chemotherapy drugs, for example, cabazitaxel, may lead to a higher rate of neutropenic events than docetaxel, however, other chemotherapy drugs were not frequently available, as discussed above. The p-values for the odds ratios were not significant, and the 95% CIs between groups were extremely wide and overlapping due to the small cohort number and the small number of events in each hormone-setting (HSPC, CRPC) (table 28). However, there was initial evidence that there may be no difference between the groups.

Both of the HSPC neutropenic events identified in the clinical noting and the trial data were identifiable in the routine data, showing some initial evidence/proof-of-concept that routine data could be used to identify events collected in the trial. Therefore, it was hypothesised that such a resource may be able to collect events not collected by the trial CRFs. The STAMPEDE trial is required to collect data on events that occurred due to the trial drug (docetaxel in this instance), however, does not mandate collection of outcomes from subsequent treatments, for example, chemotherapy administered at

diagnoses of CRPC. The HES algorithms identified events with odds ratios comparable to those derived from the note review but again due to small cohort numbers, no statistical significance was found, and thus larger cohort analyses were required. Again, the false positives that were identified in the HES data that could not be confirmed at site by note review, may have been true positives occurring at other hospitals.

This small cohort analysis enabled hypothesis generation only but suggested that routine data may be a feasible technique for collection of events not collected in a trial. It was also initial evidence that the previous hypotheses drawn from the cross-study comparisons may have been leading to biased conclusions. This was to be further investigated using different cohorts.

3.13.3 Routine data - chemotherapy

From analysing the routine HES data, in addition to the lack of clinical detail (for example, drug name), it was also identified that events were being missed (section 3.12.3.2.1). This is suspected to be due to the setting of administration; chemotherapy is normally administered in an outpatient setting and it is commonly known that historic HES outpatient coding quality was poorer, prior to accreditation (2008) (46). HSPC trial chemotherapy regimens could not be identified in the HES data (0/15 were identified). This was because of the 'Payment by Results' function of these data (section 1.2.1.1) and the funding structure for trial drugs. If the healthcare provider (for example, the hospital) has not paid for the clinical trial treatment, it would not be coded, as the hospital would not need to be reimbursed. Due to this, it was assessed that the algorithm should be developed which linked the HES data to the STAMPEDE database to accurately enable the identification of HSPC chemotherapy regimens. When this algorithm was developed, by definition, all HSPC chemotherapy regimens were identified. The majority of CRPC regimens were identified in HES, but as this is not routinely collected in the STAMPEDE trial, the HES data could not be supplemented by the trial database. Some CRPC regimens were missed due to starting another trial upon development of CRPC disease, with the same 'Payment by Results' implications and therefore lack of a routine data fingerprint. HES event quality was investigated further for neutropenic events and chemotherapy quality was also further analysed in depth in chapter 5.

The impact of these missing data on the results, needed to be assessed, so the effect of inferring missing chemotherapy regimens was undertaken. An algorithm inferring events was developed to assess the impact of this missing coding on the incidence rates. The aim was to identify if missing regimens could be inferred. A patient may experience a neutropenic event, but it may or may not be related to chemotherapy. The initial algorithm rules specified that if the neutropenic event occurred whilst not on chemotherapy, then the event was unrelated to chemotherapy and was excluded from the analysis. However, due to missing chemotherapy regimens, neutropenic events were being falsely classified as being unrelated to chemotherapy. The inferral rationale aimed to classify every neutropenic event as occurring due to chemotherapy. This was hypothesised to identify false positive chemotherapy-related neutropenic events, as it is known that neutropenic events can occur in the absence of chemotherapy. However, the aim was to identify if missing chemotherapy regimens could be identified and to assess the impact of missing chemotherapy regimens on the CRPC neutropenic event incidence.

Upon inferral, events previously classified as unrelated were reclassified as true positives, increasing the CRPC incidence (table 28). This suggested that many more neutropenic events were occurring, which could have been due to CRPC chemotherapy. A sample of these extra events that were identified by inferral were investigated by note review (six events). The note review confirmed that chemotherapy was administered in all six of these cases (table 33), the majority were clinical trial treatments, and therefore not coded for in the HES data. This suggested that the true CRPC incidence value may be greater than that without inferral and the data quality was having implications on the incidence values.

Due to the implications of these missing data, I determined that a different routine data source should be tested, that, 1) collects regimen drug details, 2) is not affected by funding structures and 3) has a potentially higher data quality. The additional routine data source I analysed was the SACT and the ability of SACT to detect chemotherapy regimens were assessed to try and enhance chemotherapy event detection and alleviate the problems identified above.

The SACT data documented chemotherapy drug details, including name and dates of administration, and therefore enabled direct analysis of docetaxel-only events. However, SACT data only began collecting in April-2012, in contrast to the HES which was available since STAMPEDE began. This led to analysis restrictions and meant that the largest cohort available for SACT analysis was 1,573 and not 3,642 participants as for the HES analyses. All patients were randomised prior to mandatory collection (July 2014) with hypothesised implications on data quality and reporting completeness. A primary function of SACT is not payment, as with the HES data. Hence, trial chemotherapy regimens should not be excluded. This SACT coding was assessed by comparing the SACT data to the HSPC STAMPEDE chemotherapy regimens (table 35). However, the quality was low due to the non-mandatory requirement to collect data in the restricted cohort (randomised 2012-2013). However, more HSPC regimens were identified than within the HES data (HES, 0%; SACT, 41.5%). SACT HSPC accuracy requires further investigation for the trial, when a more contemporary STAMPEDE cohort is available for analysis (randomised post-July 2014). As recruitment to docetaxel containing arms ceased in 2012, a much smaller sample of patients would be available, those that received docetaxel as standard-of-care. Currently too few events are estimated (discussed with the MRC CTU at UCL) to have occurred for these patients after 2014 to perform this updated analysis.

The ability of SACT to detect CRPC chemotherapy regimens was assessed by note review as these data are not routinely required to be collected in STAMPEDE. In cohort one, only five regimens occurred after SACT began collecting, but four of these were identified (all after 2014). In cohort two further investigation was undertaken; in the six patients that had missing chemotherapy events, identified upon inferral, three of these regimens were identifiable in the SACT data, after 2014. The other three occurred prior to SACT collection. In this small sample, as additional events could be identified, it is suggested that the accuracy of event detection after 2014 (when documentation was mandatory) had improved. The SACT data were therefore able to enhance the collection of CRPC events. There were also coding errors in the SACT. For example, single regimens were sometimes coded as multiple regimens and although it was possible to request various data fields (for example, dose reduction), it was clear that these fields were not

appropriate in their completeness for analyses. SACT data were provided until December-2017 as there is currently a data lag from the provider making more recent data analysis difficult. Timeliness is a further limitation of using routine data for clinical trial analysis. I have some evidence that SACT is a feasible resource to supplement trial collected CRPC therapies, but no evidence for accurate HSPC detection, due to analysing regimens prior to mandatory collection. Assessment is required in a more contemporary cohort (post July-2014).

3.13.4 Routine data - neutropenic events

Neutropenic event quality was assessed as the routine data were required to detect these events. I found that HES neutropenic event quality was improving post-2009 (figure 19). Due to these being inpatient admission events, anecdotally the early coding quality of the admitted patient care is known to be higher than that of the outpatient coding, hence the distinct accuracy values seen for inpatient events (neutropenic admissions) and outpatient events (chemotherapy administration). This is further investigated in chapter 5 by comparing surgery, an inpatient procedure with radiotherapy, an outpatient procedure. Individual neutropenic events were missing from the HES data but due to these events only being identifiable in the HES admission data, missing data could not be supplemented using additional resources. However, missing data are also present within traditional trial collected data (148, 149).

One potential cause for these missing neutropenic data were the ambiguity in the documentation and classification of neutropenic events clinically. Unified criteria are required for the correct documentation of neutropenic events. It was discussed in the NHSD coding consultation (125) that there was no simple way to classify all instances of sepsis within a single national standard, *'a single standard cannot compensate for deficiencies in the documentation, recording or coding process'*. It was also proposed that if a simplified standard was implemented, this may increase the risk of underreporting sepsis events. It was also noted that clinicians may have been reporting events as 'sepsis', where sepsis was not experienced but where infection was actually present ('infection only') (125).

Neutropenic sepsis and febrile neutropenia are often used interchangeably. This therefore, has direct implications to the routine data, as accurate administrative data relies upon distinct and accurate clinical noting. Non-standardised note completion also leads to erroneous routine data completion and subsequently missed outcomes. In the coding consultation, it was stated that *'inconsistencies and inaccuracies in the recording of sepsis within the medical record will have a negative effect on the reliability of the coded data which in turn will have a statistical and financial impact'*. There are no specific guidelines on documenting sepsis in the medical records. Thus, it was suggested that an improvement in reporting guidelines would be superior to restricting coding. The consultation made it the responsibility of the organisation to ensure that sepsis is correctly recorded for future evaluation (125). All of these analyses were undertaken using data prior to the coding consultation in 2018 (table 5), hence, no unified criteria to detect neutropenic events were available.

Due to the inconsistencies in coding and lack of standardised definitions, I set out to identify a unified definition of neutropenic events that could be identified within the restricted routine data details. Due to these restricted data, a proxy marker was created for events such as febrile neutropenia, neutropenic sepsis and infection from neutropenia - *severe neutropenic admission events*. Although not comparable to a single neutropenic definition, due to the HES coding inflexibility, this definition created a marker that could be directly comparable across the patient groups, defined as; admission for suspected severe neutropenic event, and concurrent chemotherapy administration. The routine data algorithms were therefore identifying a different neutropenic event measure, not directly comparable to previous trial definitions but it enabled direct comparison between hormone-states. The definition used to extract the relevant severe neutropenic events from the STAMPEDE data was neutropenic events grade 3 or higher, or febrile neutropenia grade 1 or higher. These should result in hospitalisation by definition and therefore is a directly comparable data source for the HES derived event definition.

There was the potential to misclassify less serious neutropenic events with more serious events or to identify miss-diagnosed neutropenic events in the routine data. Despite the requirement for an admission to exclude less severe events, patients may have been

admitted for another reason and neutropenia may have been noted as a co-morbidity. A sample of all the neutropenic event HES coding was analysed to assess this, a small proportion of events were being classified as agranulocytosis only, suggesting that patients were potentially being admitted for other diagnoses (whilst on chemotherapy) and upon which were determined to have a low neutrophil count, but not a severe neutropenic event due to chemotherapy. Therefore, to assess the distribution of severe events, the algorithm was restricted to just definitive sepsis coding (excluding agranulocytosis and neutropenia drugs band 1 only events). Upon restriction, the odds ratio showed no significant difference between groups (table 36). The HSPC neutropenic event incidence reduced 8% upon restriction and the CRPC neutropenic event incidence reduced 3%, suggesting that if HSPC chemotherapy does lead to an increased number of neutropenic events, they are potentially less severe than in the CRPC setting. Further patients could have also been admitted for a suspected event but upon admission may not have been neutropenic and this event may have still been documented due to inaccurate note completion. Anecdotally, upon admission to emergency care, patients suspected of sepsis are immediately coded as such, to enable fast treatment. The impact of this is hypothesised to have contributed to the low PPV values (75 false positives) when comparing HES detected events to the STAMPEDE trial (figure 18, A).

71% of the STAMPEDE documented neutropenic events were identified in the HES data (figure 18, A). HES identified events that were not documented as admissions in the STAMPEDE data but were documented as neutropenic events without admission in STAMPEDE. The routine data were able to detect many events; however, events were missed. Therefore, whilst not being accurate enough to act as a follow-up source alone, additional events were identified which could be used to supplement the trial data using querying techniques. When restricted to STAMPEDE admission events only, HES had an extremely high sensitivity of detection (92%) (figure 18, B). This suggested that many of the STAMPEDE events may not have led to admissions despite being documented as grade 3+ neutropenic events or febrile neutropenia and therefore would not be able to be detected in the HES admission data. When looking at accuracy of detection by year (table 37), post-2009 there was an increase in accuracy. Although a low routine data PPV was seen, the STAMPEDE trial may have underestimated the number of neutropenic

events, due to loss to follow-up, for example admission at another site (future investigation at sites is required to assess this).

Routine data may enhance trial data collection, especially if targeted data queries could be undertaken. This technique of using targeted data queries was to be investigated during this thesis (chapter 5).

3.13.5 Routine data linkage

Routine data linkage is required to enable these data to be used for clinical trial analyses, so each patient can be identified by the unique trial ID. To undertake this matching, both the trial data and the routine data are required to not contain errors in the 'linkers', for example, NHS number and DOB. As shown in the results (section 3.12.1), 227 STAMPEDE patients had errors in the NHS numbers and therefore could not be matched. Queries need to be sent to STAMPEDE sites for these 227 patients to check the invalid NHS numbers to ensure that this can be rectified for future linkage purposes. If routine data are being used for trial follow-up, during trial set up (as described in chapter 5) care should be maintained when documenting these values, or the patient will be instantly lost to follow-up via any routine data sources. The MRC CTU at UCL were informed to rectify these issues retrospectively but also to accurately document such linkers prospectively for future follow-up. This was investigated and anecdotally this occurred due to misinterpretation during telephone randomisation between the MRC CTU at UCL and the sites. I believe the MRC CTU at UCL team plan to retrospectively correct these issues and be more cautious with newly randomised patients.

The majority of the patients with valid NHS numbers were linked to the HES data (93.1%). It is to be expected that patients would interact with the NHS during the disease history (from six months prior to randomisation, to death or last follow-up) but reasons for failed linkage for the 6.9% of patients I proposed to be, 1) private healthcare, and 2) inaccurate routine data (no cancer diagnosis). HES and SACT data are not collected for non-NHS interactions and so these patients would be lost to follow-up. This would be an issue if trials are to be run solely using routine data without a technique to

query missing data. In addition, if no cancer diagnosis was present in the HES data, these data would not be provided by PHE, due to data release regulations.

The low matching rate of SACT was to be expected (49.2%), as this dataset only began collection in 2012, so any data present prior to this (2005-2012) was for retrospectively added data. Despite also restricting the linkage statistic calculation to patients randomised post 01/04/12, the rate of linkage was still exceptionally low (49.9%) (section 3.12.1). This sustained low SACT linkage rate I suspected to be due to drug type; STAMPEDE ensures all patients are at least being treated by the standard-of-care (hormone therapy) and therefore every single patient will have been administered at least one anti-cancer regimen, hormone therapy. As all anti-cancer agents are required to be documented in SACT, it would therefore be expected that all patients had at least one documented regimen in the SACT. However, it is possible that those on hormone therapy alone (arm A) may not have coded data due to the non-mandatory nature of when these data were collected. I consulted PHE regarding this who confirmed all hormone therapy prescribed within a hospital is required to be documented. However, even after the mandatory requirement for data collection in 2014, it still took time until full site conformance. Further to this, missing data may also be due to patients receiving private prescriptions, not collected in the database.

3.13.6 Algorithm rules

Although, it was identified that neutropenic events were directly identifiable in the routine data, the lack of clinical disease variables (hormone-sensitivity (HSPC, CRPC)) made the analysis challenging and *indirect* identification of outcomes was still required. The algorithms were designed to identify the incidence of events but to do this required a complex set of rules based upon many assumptions, due to the restricted clinical data in HES. Chemotherapy events and neutropenic events had to first be identified, and both had to be classified into HSPC or CRPC settings using a proxy (see section 3.11.7.1). Each event had to be confirmed as related to chemotherapy using the timing of event to enable calculation of the incidence values.

These rules could ultimately have led to misclassified events. The time of chemotherapy in relation to the date of trial randomisation was utilised as a proxy for hormone-

sensitivity. Efforts were made to reduce misclassification by undertaking the early smaller cohort analyses to confirm the ability of the model rules to identify the disease-setting correctly, by validation against note review retrieved data. Another rule that was developed included the interval where neutropenic events were hypothesised to be related to chemotherapy. The impact of widening this window to twelve, from four weeks, did not impact the overall incidence. This is suspected to be due to most chemotherapy-related neutropenic events occurring earlier in the regimen; it has been proposed that the risk may be highest following the first cycle (150). Therefore, less events may be occurring after the last cycle, regardless of the detection window.

3.13.7 HSPC vs. CRPC routine data outcomes

Cohort three (N=1,573) was chosen as the largest cohort that could utilise SACT data in addition to the HES and therefore docetaxel specifically could be investigated. Near equal incidences were identified from HSPC and CRPC docetaxel (20.5% and 20.2% respectively) with an odds ratio of 1.01, but the confidence intervals were wide, and a non-significant p-value was present. When the HES only algorithms were utilised in the largest sample (N=3,645), when restricted to sepsis-only (docetaxel was not specified), again similar incidence values were identified across chemotherapy settings. However, when analysing all potential neutropenic admissions, the CRPC event rate was marginally lower than the HSPC rate by 3.6%. Missing HES chemotherapy regimens were hypothesised to have led to this reduced CRPC rate and were further investigated by inferral. This technique (inferral), increased the CRPC rate above the HSPC neutropenic event rate. Therefore, from previous analyses, the true CRPC rate was hypothesised to lie somewhere between the HSPC upfront and CRPC relapse neutropenic event values.

The STAMPEDE data identified a HSPC neutropenic event incidence of 10%, this was lower than the HES derived neutropenic event rates (16.1% in the largest HES cohort) (table 39) but higher than the sepsis only HES-derived indicator (of 8.3%). Different events were documented in STAMPEDE and HES due to different definitions utilised by the two studies, due to the routine data coding restrictions. However, as previously discussed, it is also possible that the trial could have underestimated the number of events.

As the cohort size increased, the confidence intervals narrowed but still remained overlapping in both groups. The rates varied by method used, from between 13.3% - 20.5% of events for HSPC chemotherapy and 12.5% - 36.2% for CRPC chemotherapy. As the cohort size increased, not only the confidence intervals became narrower but the difference between the rates also converged. The impact of small single-site audits can be seen with the converging HSPC and CRPC neutropenic event incidence values. In the smallest cohort there was a large increase in CRPC neutropenic event incidence (3x), compared to HSPC. However, in the largest analyses using the same algorithm, actually a marginally lower incidence was identified for CRPC and the differences were less distinct. Hence, assumptions based upon small single-site analyses, such as seen in the literature review, should be undertaken with extreme caution. However, to reiterate, the results proposed here are also hypothesis generating and not conclusive.

3.13.8 Further considerations

There are further caveats to the patient cohorts utilised in this analysis; STAMPEDE trial patients were investigated. Trial patients inherently could be fitter than the general non-trial population and therefore these neutropenic rates may differ in the population. However, as routine healthcare data are collected automatically, there is a vast population-level data source available. If these data could be anonymised, then analysis such as the above may be possible for non-trial patients. In addition, the use of immune system stimulating drugs alongside docetaxel was also not identifiable and therefore the effect on neutropenia could not be determined.

In addition to this, patients were included that experienced chemotherapy in both the HSPC and CRPC settings. These were included in the analyses, as this reflects clinical practice. However, it was noted that a patient did experience a severe neutropenic event at both HSPC and CRPC (see section 3.12.6). Therefore, this will potentially bias the results because it is possible that the combination of HSPC and CRPC chemotherapy could influence the CRPC neutropenic events. The inclusion of this partially paired data may statistically bias the results (146). Further investigation is needed to unpick the relationship between multiple chemotherapy regimens and how this may influence the neutropenic events downstream.

3.14 Conclusion

I believe that routine data could be a feasible source for acquiring outcomes that could not have been investigated within the same patient cohort using standard trial data. Due to the inherent limitations of using routine data which were identified during the analysis, only exploratory hypothesis generation was possible. As such, no conclusive evidence could be drawn. However, routine data were used to identify an event of interest (toxicity on chemotherapy) that had not previously been investigated within a randomised cohort using standard trial data. Therefore, techniques such as this may have direct impact on patients, with the aim to enhance RCT design with the predominant purpose to increase quality and length of patient life. I identified that required events (both neutropenic and chemotherapy events) were missing and that the coding did not enable analysis of standard definitions of neutropenic events.

Routine data missingness and restricted detail have major implications when using routine data for trial analyses. Prior to using routine data for trial purposes, the scope of what data are present needs to be considered, as these data may not be available. Datasets should be rigorously validated to ensure quality and completeness for data integrity for each investigated outcome. I propose utilising routine data in addition to standard trial data as a superior method of data collection which allows detection of events not documented in the trial. Due to the data quality issues identified, chapter 5 explores an alternative method developing a framework to enhance routine data integrity, with the aim to use the routine data as the main clinical trial follow-up data source.

Despite these limitations, I have shown in this exploratory analysis that the administration of docetaxel at diagnosis for metastatic or advancing disease, may result in a similar rate of neutropenic admissions, than if given as a relapse treatment. Although not definitive evidence, first-line docetaxel should remain the standard-of-care in suitable patients with HSPC disease; the differences may be less distinct than those seen whilst making cross-study comparisons. However, further investigation is required.

This chapter aimed to assess the ability of routine data to identify a trial outcome which was *directly* available within these data (neutropenic admissions). However, it was

identified that due to the lack of data about hormone-sensitivity (HSPC, CRPC) proxies still had to be constructed. The next chapter details a study which aimed to identify outcomes not *directly* collected in the routine data, trial non-survival endpoints.

4 CHAPTER FOUR: Indirect RCT data collection (retrospective model)

4.1 Disclosure

I presented work from this chapter at the *Trials using Cohorts and Routine Health Data international symposium* on the 15th May 2019 and the abstract is subsequently planned to be published in *Trials* journal (in press). I have also been accepted to present work from this chapter at the International Clinical Trials Methodology Conference (ICTMC) in October 2019. All ICTMC abstracts are proposed to be published within *Trials* journal.

4.2 Abstract

4.2.1 Background

Robust, validated trial endpoints (surrogates) that occur earlier than overall survival (OS) are desirable to expedite oncology trials, as long follow-up is often required to reach the OS endpoint. Routine administrative healthcare data have the potential to supplement standard trial conduct; however, such databases often contain limited and erroneous data, meaning standard trial ‘time to event’ analysis is not directly possible. I proposed major outcomes could be *indirectly* identified in routine datasets and demonstrate a worked example of development and validation of a novel routine data trial surrogate. This work was embedded in the STAMPEDE trial using the Hospital Episode Statistics (HES).

4.2.2 Methods

A subset of STAMPEDE patients had details of their hospital service interactions extracted and triangulated from three datasets: STAMPEDE trial data, clinical records and routine data records. An algorithm was developed with the aim to process routine healthcare data, to identify events such as progression, capable of predicting OS differences. Hazard ratios (HR) were calculated using Cox-proportional hazards models to compare treatment effects and were compared to standard STAMPEDE non-survival endpoints. Correlation analyses were conducted to determine the strength of association of the routine data-derived endpoint to traditional trial endpoints.

4.2.3 Results

To identify outcomes of interest, the algorithm was based on prostate cancer activity (healthcare interactions) over defined time periods of 8-weeks. Although the algorithm was not identifying events directly analogous to any single standard trial endpoint, such as failure, metastases or progression-free survival (FFS, MFS, PFS), the model identified a composite, which had elements of all three. During initial validation (N=46), 36/46 patients experienced a FFS, MFS or PFS event and of which, the Hospital Episode Statistics (HES) algorithm corresponded to 30/36 (14/33 FFS, 23/29 PFS and 22/28 MFS). HRs capable of identifying treatment benefits were found, comparable to trial data methods. For example, in the largest validation cohort (N = 1,695), the routine data HR was 0.88 (95% CI 0.77 – 1.01), compared with the STAMPEDE MFS HR of 0.82 (95% CI 0.71 – 0.95).

4.2.4 Conclusion

Despite not being able to identify standard trial non-survival endpoints in routine healthcare data, I developed a novel oncology endpoint based upon HES-activity – ‘activity-free survival’ (AFS). This enabled the estimation of OS treatment effects and identified events unreported by the trial. However, prior to use in other settings, validation is essential. Such a technique is proposed to reduce clinical burden, time, resources and costs. In addition, this technique could be used to identify and validate events missed in a trial and could enable comparison of trial data to real-world data. This activity-based indicator may also have the added benefit of directly correlating with patient quality-of-life.

4.3 Introduction

When patients are randomised into randomised controlled trials (RCT), follow-up is required for documentation of events that have occurred, and these are subsequently incorporated into the trial database. Follow-up reports events, such as, when patients have had a change in their disease: progression, metastases, or failure. These disease outcomes are required to enable treatment efficacy comparisons to be made. Overall-survival (death) is often the most 'valued' endpoint for researchers and patients in oncology trials, as it is unambiguous, however, this can often take a long time to reach. Therefore, these other surrogate events that are collected at follow-up, can be used to determine treatment efficacy, prior to follow-up for OS being met. New and validated oncology trial endpoints are highly sought after, to *'provide a robust surrogate for OS that will expedite the design and conduct of future adjuvant therapy trials'* (151). The overall aim is to increase the quality and length of patient life. In addition to this, patient reported outcome measures (PROM) are also central within clinical trials (152) and these outcomes are dictated around the patient experience; for example, quality of life assessments. It has been proposed that the *'greatest potential benefit for the majority of patients, whether survivors or not, will be through addressing quality of life issues'*.

Documenting trial outcomes through standard patient-clinician follow-up techniques can be burdensome for patients (153, 154) and clinical staff, has high costs associated with it, and uses a large amount of resources and time (154). Loss to follow-up is also a known trial limitation where events may go unreported (66, 155). The main aim of this thesis (section 1.8.1) was to investigate if routinely collected healthcare data could identify trial outcomes in oncology; hence, the ability to detect surrogate endpoints using routine data was to be investigated. Routine data may offer opportunities for endpoint development and long-term tracking of patient response, however, major challenges are acknowledged (2).

As discussed in the introduction (1.2.1), healthcare systems collect extensive administrative patient records, for example the Hospital Episode Statistics (HES) data (27). This documents National Health Service (NHS) inpatient, outpatient and emergency visits, containing diagnoses and procedure records (populated by International Statistical Classification of Diseases (ICD) (32) and Office of Population Censuses and

Surveys Classification of Interventions and Procedures (OPCS) codes (33) respectively). As explored in chapter 3 and chapter 5, if the event is identifiable in the routine data coding, such as sepsis admissions and surgery, these events can be *directly* identified. However, the HES contain no data on disease or treatment outcomes (patient reported or non-patient reported outcomes) and therefore, major trial endpoints, for example, progression-free survival (PFS) cannot be *directly* found. Hence, I proposed that routine administrative data could *indirectly* be used to identify these outcomes.

To my knowledge, worldwide, there is limited evidence for using routine healthcare records to *indirectly* identify oncology outcomes and extremely limited evidence of developing such models in a trial setting. I also believe that a novel routine data surrogate proxy with the ability to replicate results produced from standard oncology clinical trial outcomes has never been developed, nor validated; including, no studies have been published using HES data to identify oncology trial outcomes. Hence, a validated method to identify clinically useable trial surrogates in oncology from routine data is overdue. This study was embedded in the STAMPEDE trial (introduction, 1.5.1). The primary outcome of STAMPEDE is overall survival (OS) and secondary endpoints include PFS, metastases-free survival (MFS), failure free-survival (FFS) and skeletal related events (SRE) (see introduction, 1.5.3 for details); STAMPEDE also collects quality of life data (72).

The aim of this chapter was to generate a validated endpoint using routinely collected data. Routine data are frequently used to identify survival-based endpoints (156) due to mortality records being collected *directly* in these data. However, non-survival based endpoints are infrequently derived from routine data (despite the known potential) due to the limitations of the data source; hence, this being the motivation for the study.

Here I present a systematic literature review, followed by the outline of a novel method developed to extract trial outcomes from routine healthcare records. This is presented in three stages; model development, algorithm training and algorithm validation.

4.4 Systematic literature review: Abstract

4.4.1 Background and objectives

This systematic review aimed to identify if methods had been developed to utilise routine data to identify clinical trial non-survival endpoints or events (for example, recurrence and progression) in oncology.

4.4.2 Search methods

A systematic review of literature was undertaken across four electronic databases: Web of Science (WoS), PubMed, MEDLINE and EMBASE, plus manual reference checks. The studies were screened against inclusion and exclusion criteria. Content analysis was undertaken to assess weakness and bias and a thematic summary was also performed to identify method themes.

4.4.3 Results

A total of 661 articles were extracted from the literature systematic search, with nine further identified from searching the included full-text reference lists. 50 papers remained after duplicates, unrelated titles and conference abstracts were excluded and 37 remained after the abstracts were assessed for eligibility. 27 studies, which were identifying subsequent oncology events using routine data, were included for the analyses after the full texts were screened. No papers were identified which developed, validated and utilised a clinically relevant indirect methodology for RCT purposes in the UK; Worldwide, two studies were identified.

4.4.4 Authors conclusions and implications

A method to extract oncology events from trial data would uncover hidden but routinely available, cheap, rich outcome data to allow further clinical questions to be answered, enhancing clinical trials and allowing more expedient development of new therapies.

4.5 Systematic literature review: Aims and objectives

This literature review aimed to assess the current methods used to extract outcome events from routine data sets. The objective of an initial search was to identify studies where routine data were linked to RCTs; however, only two studies were found to undertake this, so the search was broadened to trial and non-trial methodologies (identifying trial-related outcomes) which are presented below.

4.6 Systematic literature review: Materials and methods

4.6.1 Criteria for selecting studies for this review

4.6.1.1 Review question

What models have been developed to utilise routine healthcare data to identify oncology clinical trial-related outcomes (not including those *directly* using metastases diagnoses codes only)?

4.6.1.2 Types of studies

Included research designs were: single case reports to RCTs documenting the use of a methodology, using routine data for the analyses. Systematic reviews of meta-analyses were excluded; however, the references were analysed for inclusion of the primary resources.

4.6.1.3 Types of participants

Any oncology patients whose routine healthcare data had been analysed.

4.6.1.4 Types of interventions

Creation of an algorithm to identify clinical trial-related disease events, including, recurrent, progressive and metastatic disease markers.

4.6.1.5 Types of outcome measures

Algorithm design/method to identify clinical trial-related disease events including, recurrent, progressive, and metastatic disease markers.

4.6.2 Search methods for identification of studies

4.6.2.1 Electronic searches

The search was undertaken using four databases to 31/05/19; WoS (157), PubMed (158), Medline (159) and Embase (159). The final search for the whole review period was re-run on the 31/05/19, despite the first search being undertaken at the beginning of November 2016.

The search was not restricted by date or country as it was important to gain an international perspective. The search strategy used can be seen broadly in 4.6.2.2 and the details can be seen in table 40. I carried out all of the searches, however, Samantha Johnson from the University of Warwick library gave assistance, where required.

4.6.2.2 Main constituents of literature search

- Database: Hospital episode statistics, routine, administrative, registry, claims data
- Disease: Neoplasm, cancer
- Intervention: Algorithm
- Outcome: Recurrent disease, progressive disease, metastatic disease

Database	Search criteria	
Web of Science (WoS)	TS=((hospital episode statistic* OR routine NEAR/4 data OR administrative NEAR/4 data OR registry OR registries OR claims NEAR/4 data) AND (recurrence OR progressive disease OR progression OR metastases OR metastatic) AND (algorithm) AND (neoplasm OR cancer))	
PubMed	((hospital episode statistic* OR routine NEAR/4 data OR administrative NEAR/4 data OR registry OR registries OR claims NEAR/4 data) AND (recurrence OR progressive disease OR progression OR metastases OR metastatic) AND (algorithm) AND (neoplasm OR cancer))	
MEDLINE & Embase	1	hospital episode statistic*.mp. [mp=ti, ab, hw, tn, ot, dm, mf, dv, kw, fs, nm, kf, px, rx, ui
	2	(routine adj3 data).mp. [mp=ti, ab, hw, tn, ot, dm, mf, dv, kw, fs, nm, kf, px, rx, ui
	3	(administrative adj4 data).mp. [mp=ti, ab, hw, tn, ot, dm, mf, dv, kw, fs, nm, kf, px, rx, ui
	4	registry.mp. [mp=ti, ab, hw, tn, ot, dm, mf, dv, kw, fs, nm, kf, px, rx, ui
	5	registries.mp. [mp=ti, ab, hw, tn, ot, dm, mf, dv, kw, fs, nm, kf, px, rx, ui
	6	(claims adj4 data).mp. [mp=ti, ab, hw, tn, ot, dm, mf, dv, kw, fs, nm, kf, px, rx, ui
	7	recurrence.mp. [mp=ti, ab, hw, tn, ot, dm, mf, dv, kw, fs, nm, kf, px, rx, ui
	8	progressive disease.mp. [mp=ti, ab, hw, tn, ot, dm, mf, dv, kw, fs, nm, kf, px, rx, ui
	9	progression.mp [mp=ti, ab, hw, tn, ot, dm, mf, dv, kw, fs, nm, kf, px, rx, ui
	10	metastases.mp. [mp=ti, ab, hw, tn, ot, dm, mf, dv, kw, fs, nm, kf, px, rx, ui
	11	metastatic.mp. [mp=ti, ab, hw, tn, ot, dm, mf, dv, kw, fs, nm, kf, px, rx, ui
	12	algorithm*.mp. [mp=ti, ab, hw, tn, ot, dm, mf, dv, kw, fs, nm, kf, px, rx, ui
	13	1 or 2 or 3 or 4 or 5 or 6
	14	7 or 8 or 9 or 10 or 11
	15	12 and 13 and 14
	16	neoplasm*.mp. [mp=ti, ab, hw, tn, ot, dm, mf, dv, kw, fs, nm, kf, px, rx, ui
	17	cancer.mp. [mp=ti, ab, hw, tn, ot, dm, mf, dv, kw, fs, nm, kf, px, rx, ui
	18	16 or 17
	19	15 and 18
	20	remove duplicates from 19

Table 40: The systematic search terms used to conduct this literature review.

4.6.2.3 Searching other resources

To identify missed eligible publications, reference lists were analysed, and the relevant articles were screened.

4.6.3 Exclusion and inclusion criteria

Studies were excluded and included to identify the desired 1) type of literature and 2) methods utilised to identify the outcomes. The inclusion and exclusion criteria are shown in table 41.

Inclusion criteria	Exclusion criteria	Reason
Types of literature		
-	Abstract only or grey literature, including theses and posters	Peer-reviewed full-texts were required to ensure validated details could be extracted for the review
-	Studies not translated into English	To ensure details could be extracted
-	Summary/review papers	To ensure details could be extracted. The primary studies in the summary/review papers were however, screened for inclusion
Methods utilised		
-	Direct identification of outcomes (unless, RCT linkage was present)	<p>Administrative databases do not routinely collect data on outcomes (<i>directly</i>); events therefore must be inferred (<i>indirectly</i>)</p> <p><i>Directly</i>: analysing diagnosis coding. The recording of such variables is not mandatory, so the coding is limited. Therefore, this technique was excluded, as is not a feasible method RCT data collection.</p> <p><i>Indirectly</i>: analysing clinical coding indicators (focus of the review)</p> <p>The aim was to find papers identifying clinical trial outcomes, therefore, to ensure all papers were captured, if the research involved an RCT but only used <i>direct</i> coding, then the paper could be included. Otherwise, as stated, the use of <i>direct</i> coding only was an exclusion criterion</p>
Oncology outcomes of interest e.g. progression, metastases, recurrence	-	The focus of this chapter was to identify oncology outcomes not routinely collected in administrative data. RCTs already collect events e.g. incident cases and diagnosis stage and hence these are not outcomes desired to be collected here
Papers developing algorithms to identify the outcomes of interest	-	The methods designed to extract the outcomes were a pivotal part of the review and hence papers were required to document the development of algorithms. Those validating previous algorithms were excluded
Outcomes detected using routine administrative databases	-	Administrative data use was required; non-routine, manually collected databases with limited RCT feasibility, for example, registries and laboratory data, were excluded
-	Papers developing predictive models for outcomes	Interventional RCTs are run to identify response to treatment, hence, predictive outcomes were not of such interest for this study

Table 41: The inclusion and exclusion criteria for the review.

4.6.4 Data collection process

The search included methods as described in the Cochrane Handbook for Systematic Review of Interventions (136). After the search articles were generated using the strategy (table 40), references from the four databases were manually screened, and where possible automatically screened, for duplicates. Endnote (108), the reference manager, was used to collect and report the studies. For assessment in the review,

firstly, titles were screened, secondly, abstracts were screened and finally the full texts were screened; articles were removed throughout this process. From abstract screening onwards, information was extracted for included publications and reasons for exclusion were coded. These data were extracted from the literature using predesigned Microsoft Excel data collection forms. Data items were also collected for the content weakness and bias analyses in a Microsoft Excel pre-defined data collection table. 17 criteria were developed which were influenced by the criteria within the Critical Appraisals Skills Programme (CASP) guidelines (160). The criteria were analysed to homogenise different aspects of the literature. The criteria included: participant, intervention/algorithm, outcome and overall quality assessments.

Data were collected for the thematic analyses in a Microsoft Excel document; however, the content of the studies informed the design of the thematic groups and hence, the data collection tool was retrospectively designed. The studies were also analysed in depth to identify the main methodological constituents of the algorithms and the relevant details were extracted.

4.6.5 Reporting

A narrative synthesis qualitative systematic review was undertaken. The PRISMA guidelines (135) influenced (see appendix 8.2.1) the reporting of the systematic search and the PRISMA flow diagram was used to graphically present the results (135). Content analyses (to assess weakness and bias) and a thematic summary (to assess the methods designed to identify outcomes) were reported (161). In addition, the different algorithmic methods identified were reported in individual flow diagrams.

4.6.6 Assessment of heterogeneity

A multiple of outcomes and content were analysed with methodological heterogeneity, hence combining them would have led to clinically meaningless interpretation. Thus, due to the literature assessing non-comparable quantities, a meta-analysis was not undertaken (162). Alternative methods were adopted to assess heterogeneity and relationships between studies; the content and thematic analyses (161).

4.6.7 Risk of bias

A custom risk of bias assessment was undertaken during content analyses. This consisted of assessment at a participant, intervention/algorithm, outcome and overall quality level. The tool that was developed to undertake these analyses can be seen in the appendix (section 8.2.3).

4.7 Systematic literature review: Results and discussion

4.7.1 Study selection - PRISMA

The PRISMA flow diagram (figure 21) (135), shows the various stages undertaken during the review. At the end of the four database searches, 27 papers were included in the review.

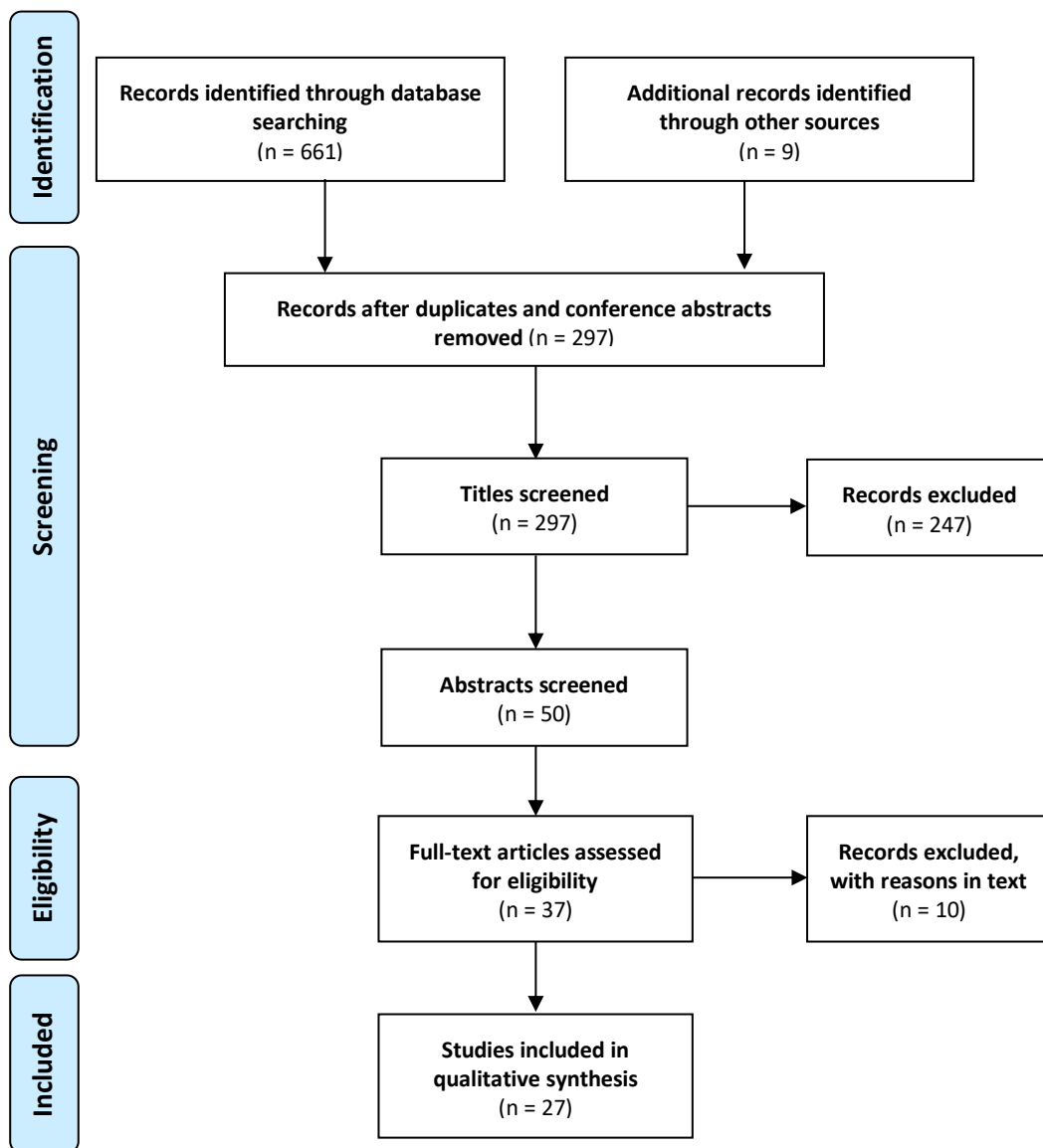


Figure 21: A PRISMA flow diagram showing the stages of the review.

4.7.2 Summary of the literature identified

Summary of findings tables can be seen in table 42 and table 43, showing aggregated data from the included studies.

Author	Date first available (E = EPub)	Journal	Title	Country of routine database	Oncology setting	Outcome of interest (paper term)	Routine data databases to identify subsequent event (excluding reference to validate or identify pts)
Earle (163)	2002	Medical care	Identifying cancer relapse using SEER-Medicare data	USA	Leukaemia	Recurrence	Medicare
McClish (164)	2003	Journal of clinical epidemiology	Using Medicare claims to identify second primary cancers and recurrences in order to supplement a cancer registry	USA	Mixture	Recurrence & 2nd primary	Medicare
Lamont (165)	2006	Journal of the National Cancer Institute	Measuring disease-free survival and cancer relapse using Medicare claims from CALGB breast cancer trial participants	USA	Breast	Recurrence	Medicare
Anaya (166)	2011 (E), 2012	Journal of Surgical Research	Use of administrative data to identify colorectal liver metastasis	USA	Colorectal	Metastases	Veterans Administration (VA) administrative system
Chubak (167)	2012	Journal of the National Cancer Institute	Administrative data algorithms to identify second breast cancer events following early-stage invasive breast cancer	USA	Breast	Recurrence & 2nd primary	± Surveillance, Epidemiology, and End Results (SEER) data Group Health administrative databases
Kimmick (168)	2012	Journal of geriatric oncology	Adjuvant Radiation and Outcomes After Breast Conserving Surgery in Publicly Insured Patient	USA	Breast	Recurrence	SEER-Medicare
Nordstrom (169)	2012	Pharmacoepidemiology and Drug Safety	Identification of metastatic cancer in claims data	USA	Mixture	Metastases	Medical & pharmacy claims data
Liu (47)	2013	Clinical Otolaryngology	Using routine data to estimate survival and recurrence in head and neck cancer: our preliminary experience in twenty patients	UK	Head & neck	Recurrence	HES SACT RTDS DBS
Hagberg (170)	2013	Cancer epidemiology	Incidence of bone metastases in breast cancer patients in the United Kingdom: Results of a multi-database linkage study using the general practice research database.	UK	Breast	Metastases	General Practice Research Database (GPRD) HES National Cancer Registry (NCR)
Lash (171)	2014 (E), 2015	International Journal of Cancer	A validated algorithm to ascertain colorectal cancer recurrence using registry resources in Denmark	Denmark	Colorectal	Recurrence	Danish National Registry of Patients (treatment data) Danish Civil Registration System Danish Pathology Registry (DPR)

Ricketts (172)	2014	Computer Methods and Programs in Biomedicine	Automated estimation of disease recurrence in head and neck cancer using routine healthcare data	UK	Head & neck	Recurrence	HES SACT RTDS DBS
Ehrenstein (173)	2015	Clinical epidemiology	Validation of algorithms to detect distant metastases in men with prostate cancer using routine registry data in Denmark	Denmark	Prostate	Metastases	Danish National Registry of Patients Clinical Laboratory Information System (LABKA research database)
Deshpande (174)	2015	Annals of epidemiology	Development of a claims based algorithm to identify colorectal cancer recurrence	USA	Colorectal	Recurrence	Inpatient and outpatient hospital billing data
Haque (175)	2015	Medical care	A hybrid approach to identify subsequent breast cancer using pathology and automated health information data	USA	Breast	Recurrence & 2nd primary	Clinical (pathology reports - Only used automated algorithm if pathology report lacked information) and administrative database
Warren (176)	2016	Medical care	Sensitivity of Medicare Claims to Identify Cancer Recurrence in Elderly Colorectal and Breast Cancer Patients	USA	Mixture	Recurrence	SEER (surgery only) Medicare
Livaudais (177)	2016	Generating evidence and methods to improve patient outcome	A Validation Study of Administrative Claims Data to Measure Ovarian Cancer Recurrence and Secondary Debulking Surgery	USA	Ovarian	Recurrence	Claims data
Nordstrom (178)	2016	Frontiers in oncology	Validation of claims algorithms for progression to metastatic cancer in patients with breast, non-small cell lung, and colorectal cancer	USA	Mixture	Metastases	Claims data
Joshy (179)	2016 (E), 2017	Asia-Pacific Journal of Clinical Oncology	Validating a proxy for disease progression in metastatic cancer patients using prescribing and dispensing data	Australia	Mixture	Metastases	Prescribing and dispensing data
Hassett (180)	2017	Medical care	Detecting Lung and Colorectal Cancer Recurrence Using Structured Clinical/Administrative Data to Enable Outcomes Research and Population Health Management	USA	Mixture	Recurrence	Medicare (claims) Cancer Research Network Virtual Data Warehouse (CRN-VDW) (claims)
Gupta (181)	2017 (on paper), 2018 (E)	Medical care	Validity of Administrative Data in Identifying Cancer-related Events in Adolescents and Young Adults: A Population-based Study Using the IMPACT Cohort	Canada	Cancers in adolescents - mixture	Subsequent cancer events (progression, relapse,	Discharge database health insurance database ambulatory care database

							second cancer)
Ritzwoller (182)	2017 E, 2018	Journal of the National Cancer Institute	Development, Validation, and Dissemination of a Breast Cancer Recurrence Detection and Timing Informatics Algorithm	USA	Breast	Recurrence	Cancer research network (CRN) (containing EHR, administrative/claims and registry data)
Wong (183)	2018	BMC health services research	A methodology to extract outcomes from routine healthcare data for patients with locally advanced non-small cell lung cancer	UK	Lung	Recurrence	HES SACT RTDS personal demographics service (PDS)
Rasmussen (184)	2018	Clinical epidemiology	A validated algorithm to identify recurrence of bladder cancer: a register-based study in Denmark	Denmark	Bladder	Recurrence	Various Danish databases
Uno (185)	2018	JCO clinical cancer informatics	Determining the time of cancer recurrence using claims or electronic medical record data	USA	Mixture	Recurrence	Medicare CRN electronic medical record and claims (virtual data warehouse)
Xu (186)	2019	Head and neck	Developing case-finding algorithms for second events of oropharyngeal cancer using administrative data: A population-based validation study	Canada	Head & neck	Second events	Claims data discharge abstract data national ambulatory care reporting system cancer registry vital statistics cancer measurement outcomes research and evaluation unit data
Xu (187)	2019	BMC cancer	Development and validation of case-finding algorithms for recurrence of breast cancer using routinely collected administrative data	Canada	Breast	Recurrence	Claims data discharge abstract data national ambulatory care reporting system cancer registry vital statistics cancer measurement outcomes research and evaluation unit data
Rasmussen (188)	2019	Cancer epidemiology	A validated algorithm for register-based identification of patients with recurrence of breast cancer-Based on Danish Breast Cancer Group (DBCg) data	Denmark	Breast	Recurrence	Various Danish databases (four national registers)

Table 42: Part 1) Summary of findings table for the 27 papers identified.

Author	Total patient no./event no.*	Method validation	Event timing found?	Timing of recurrence notes	Summary statistics: Sensitivity, specificity, PPV, NPV (ranges where >1 algorithm were developed)					Successful? (>80% accuracy of all available summary statistics)
Earle (163)	89/22	Medical record review	Y	Median = 6.5d after the medical record	0.86 - 0.91	0.91 - 0.99	0.77 - 0.95	0.96 - 0.97		(+)
McClish (164)	15043/335 (recurrence)/321 (2nd primary) (Validation: 45/23)	Registry plus small record review	N	NA	N/S	N/S	0.51 - 0.58	N/S		(-)
Lamont (165)	45/12	Trial data	Y	Disease-free-survival analysis completed. Median = 40d & IQR = 12-86d. Mean difference = 99d	0.83 - 1.00	0.95 - 0.97	N/S	N/S		(+)
Anaya (166)	924/302	Medical record review	N	NA	0.69	0.97	0.87	0.92		(-)
Chubak (167)	3152/299 (recurrence)/93 (new prim)/15 (both)	Medical record review	N	NA	0.69-0.99	0.81-0.99	0.37-0.90	0.97-1.00		(+)
Kimmick (168)	416/53 (of the 60 identified in the routine data)	Medical record review	N	NA	0.88	0.9	N/S	NS		(+)
Nordstrom (169)	3415/1043	Oncology EHR data warehouse	N	NA	0.60 - 0.81	0.75 - 0.97	0.75 - 0.86	0.67 - 0.95		(-)
Liu (47)	20/11	Medical record review	Y	Survival curves created. 16/20 pts = 'acceptable agreement' (to the PFS); Kendall's Tau: 0.78	0.91	N/S	N/S	N/S		(+)
Hagberg (170)	100/100 (chose pts identified with event in the routine data, GPRD)	GP questionnaires	N	NA	NS	NS	NS	NS		NS
Lash (171)	335 (actively followed cohort)/63	(actively followed cohort) = trial data (the 335	Y	Plotted cumulative incidence of recurrence. The cumulative incidence 'overlapped substantially'	0.95	0.97	0.86	0.99		(+)

observational study validation sample)										
Ricketts (172)	122/40	Medical record review	Y	4/21 correct date; 7/21 in <7days; 9 in <1m - implied 1 further patient >1m; Kendall's Tau: 0.82	0.33 - 0.53	0.87 - 0.98	N/S	N/S	N/S	(-)
Ehrenstein (173)	212/43	Medical record review	Y	Noted whether the true outcome occurred in a 30, 90, 180 or within any date period. 11/73 events confirmed for all follow-up	N/S	N/S	0.03 - 0.28	0.97 - 1.0		(-)
Deshpande (174)	174/32	Medical record review	N	NA	0.06 - 0.13	0.99 - 1.00	0.67 - 1.00	0.83		(-)
Haque (175)	175/62 (train set), 500/65 (test set)	Medical record review	Y	Date differed by <60d from the medical record in >90% of patients	0.97	0.92	0.66	1.00		(-)
Warren (176)	10736 (no recurrence reference)	None	Y	No validation was undertaken	NS	NS	NS	NS		NS
Livaudais (177)	94/32	Medical record review	N	NA	1.00	0.89	N/S	N/S		(+)
Nordstrom (178)	1017/111	Cancer registry with supplemental manual chart review	N	NA	0.47 - 0.62	0.73 - 0.96	0.19 - 0.45	0.85 - 0.98		(-)
Joshy (179)	171 (validation)/90	Medical record review	N	NA	0.64 - 0.74	0.81 - 0.88	0.56 - 0.61	0.91 - 1.00		(-)
Hassett (180)	4527/744	Medical record review or registry (cancer care outcomes research and surveillance study (CanCORS))	Y	recurrences classified <6m= 81% lung; 85% colorectal CanCORS/Medicare; 75% lung; 71% colorectal CRN-VDW	0.72 - 0.77	0.94 - 0.98	0.82 - 0.88	0.91 - 0.97		(-)
Gupta (181)	1404/354	Medical record review	Y	Median difference to note review = 27d; interquartile range = 11.5-54d; difference <30d in 49.2% of cases & <60d in 74.3%	0.56	0.92	0.71	0.86		(-) (without chart to identify the end of initial treatment)
Ritzwoller (182)	7331/542	Medical record review	Y	(Training set) 23.4% & 67.3% of recurrences correctly classified in 3 & 6m; (validation set) 20.1% (3m) & 68.6% (6m) and 14.3% (3m) & 64.7% (6m) in the two validation sets at 3 & 6m	Rec status: 0.70 - 0.81	Rec status: 0.97 - 1.00	Rec status: 0.66 - 0.91	Rec status: 0.98 - 0.99		(-)

Wong (183)	43/16	Medical record review	Y	4/16 identified <2w of the reference (incl. 34 exactly); 6/16 >4w but <100d; 2/16 >100d	0.75	1	N/S	N/S	(-)
Rasmussen (184)	187/128	Danish bladder cancer database	Y	The recurrence was estimated <30d of the reference for 64%; <90d in 90%	0.85	0.9	0.95	0.74	(-)
Uno (185)	(recurrence status) 308/89, (timing of recurrence) 600/84, 792/84, 2827/355	Medical record review from CanCORS CRN tumour registry (virtual data warehouse)	Y	Average prediction error = 4.8m, 4.8m, 4.9m & 5.4m for the four different datasets/diseases	NS	NS	NS	NS	NS
Xu (186)	568/124	Medical record review	N	NA	0.52 - 0.88 - 0.88 -	0.85 - 0.99	0.61 - 0.94	0.88 - 0.96	(-) (not incl. algorithm using note review to confirm 109 discordant cases)
Xu (187)	598/121	Medical record review	N	NA	0.75 - 0.94	0.93 - 0.98	0.79 - 0.93	0.94 - 0.99	(+)
Rasmussen (188)	471/149	Danish breast cancer group registry (including validation in the Danish national pathology register during a previous study) GP review small medical record review if no event was identified	Y	Median difference to the reference = 17d; IQR 6-56d	0.97	0.97	0.94	0.99	(+)

Table 43: Part 2) Summary of findings table for the 27 papers identified.

* = where possible the numbers quoted are for when events were compared to the reference; m = months; w = weeks.

Papers were published over a range of years, from 2002 until 2019 (end of May). There was an overall increase in the number of papers being published over time (figure 22) (table 42).

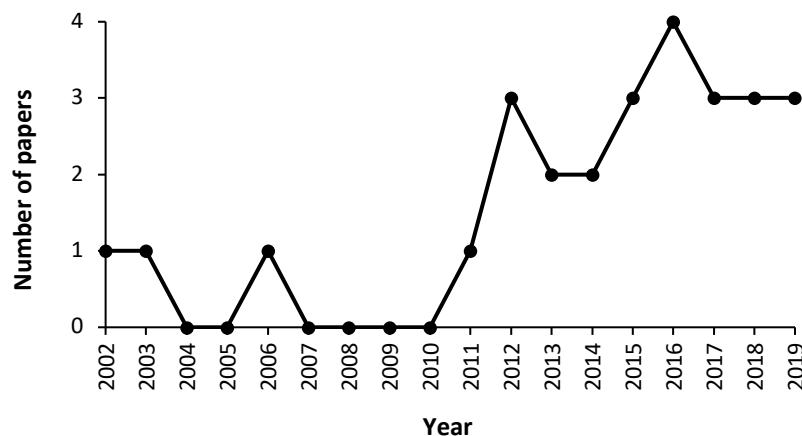


Figure 22: The number of papers published between 2002 and 2019 (until May).

The majority of databases used were in the USA (15/27, 56%), however others were Denmark (4/27, 15%), United Kingdom (4/27, 15%), Canada (3/27, 11%) and Australia (1/27, 4%) (table 42). Research was undertaken in many different fields of oncology: leukaemia, breast, colorectal, head and neck, prostate, ovarian, lung, bladder and mixtures (including, cancers in adolescents). The most common field was breast (8/27, 30%) with only one study (1/27, 4%) investigating bladder and prostate cancer individually, both in Denmark (table 42). Various subsequent cancer events were also identified; recurrence, second primary (breast cancer), metastases and progression (table 42). Often subsequent event terms were used interchangeably, but the overall aim for all studies was to identify subsequent cancer events.

Only four papers (4/27, 15%) utilised the Hospital Episode Statistics (HES) data and identified outcomes across three specialities, head and neck, breast and lung cancer. Three of these studies were from the same research group (47, 172, 183) and one was from a different group which also used the general practice research database (GPRD) (now known as the CPRD) (170). In all four studies, HES data were used in addition to further databases; Systemic Anti-Cancer Therapy Dataset (SACT), National Radiotherapy

Dataset (RTDS), Demographic Batch Service (DBS), Personal Demographics Service (PDS), the GPRD and the National Cancer Registry (NCR) (table 42).

Other databases within the other studies, included: Medicare (a US claims database), the Veterans Administration (VA) data and various Danish registries (table 42). The total number of databases used for each study varied from one to six. Utilising more databases enables more variables to be analysed, however, the time and resources needed to acquire and link the data sets increases and thus reduces the generalisability. The same principle applies for the type of coding utilised in the databases. If ICD codes are utilised (used in the majority of papers), the algorithms are more feasible for use internationally. However, often national coding systems were utilised, for example, in Denmark, restricting the applicability of the algorithm internationally, further reducing generalisability.

The studies varied in size; the total cohort ranged from N=20 (47) to N=15,043 patients (164) and the number of subsequent cancer events validated ranged from N=11 (47) to N=1,043 events (169). Validation was undertaken using various resources, the most common being medical record review, but other data sources for validation included: registries, electronic medical record warehouses (EMRs), general practice (GP) questionnaires, RCT data, and one study did not validate (176). Where medical note review was undertaken for validation, the largest number of events validated were 542 (182). Only two studies validated the algorithms using RCT data (165, 171).

Fifteen studies identified the timing of the event, hence, the remaining twelve only identified if a patient had experienced an event over the whole follow-up period, regardless of the date. There were varying levels of accuracy of event detection; using the Warren and Yabroff (189) described cut-off of administrative data event accuracy (determined as a sensitivity, specificity and positive predictive value of greater than 80%), only 9/27 (33%) had a positive outcome (including the further requirement for a negative predictive value of greater than 80%). 15/27 (56%) studies documented poorer accuracies and three studies did not publish descriptive statistics.

4.7.3 Content analysis – weakness and bias

The participant, intervention/algorithm, outcome, and quality criteria for these studies were assessed during the content analyses. This was an opportunity to critically appraise the literature for sources of weakness and bias (appendix, 8.2.2, table 79, table 80).

4.7.3.1 Participant weakness and bias

The criteria studied are discussed below:

One: Accurate identification of the study cohort

Correctly identifying the patient cohort that are known to have a disease is vital for validation of the algorithms to detect events. Where cohorts were identified from varied across the studies. The majority of papers (24/27, 89%) (figure 23) identified cohorts from a validated source; a clinical trial cohort, a cancer registry or acquired from note review.

However, two papers identified patients using only routine data (166, 177). Anaya (166) identified cases using an administrative diagnosis of colorectal cancer and thus, as stated in the discussion, may have led to '*a systematic misclassification bias of potential cases*'. Livaudais (177) similarly identified those diagnosed with ovarian cancer from the administrative data warehouse coding, again potentially leading to misclassification of patients and bias. It was also not specified where the Xu (186) cohort was identified from.

Due to not having a 'true' gold standard to compare to, I am not able to confirm if eligible patients were included or not. Hence, these three algorithms may have been built using ineligible patients, potentially impacting the ability to identify events in a cohort of 'true' eligible patients. This may have led to inaccurate measures of event detection sensitivity and as such a true measure of algorithm performance may not be possible.

For example, if the algorithm was built to detect cancer events in patients without cancer (due to inaccurate identification of the cohort), when the algorithm is exposed to

data for a patient with cancer, 'true' events may be missed. Therefore, false negatives and positives may be present using the 'truly' eligible cohort, biasing the results.

Two: Event sample size for model development or validation suitable

A priori, a cohort number greater than or equal to 50 was deemed an appropriate sample size by myself and Dr Helen Parsons. The value of 50 was chosen as this was hypothesised to allow a reasonable sample to be confirmed. This figure has been previously regarded as the minimum to estimate precise measures of the standard deviation. Sim et al propose that for a high level of statistical confidence '*a pilot study of at least n=50 is advisable*' (190).

Although overall cohort numbers were often seen to be large, the number of subsequent events to be validated was the most important value. For example, one paper analysed only 11 events across 20 patients (47), whilst another analysed a larger cohort of 212 where only 43 patients experienced an event (173). Likewise, in the largest study (164), although over 15,053 patients were in the cohort, only 2% (335) experienced a recurrence event. The majority of papers, 21/27 (78%) (figure 23) analysed greater than or equal, 50 events.

The six papers that investigated less than 50 events included, Earle (22 events) (163), Lamont (12 events) (165), Liu (11 events) (47), Deshpande (32 events) (174), Livaudais (32 events) (177) and Wong (16 events) (183).

The larger the sample size, the more likely it is to find a range of patient disease pathways or trajectories. Hence, the more generalisable an algorithm may be. Conversely, if the algorithm has been built in a small sample, it is possible that the algorithms developed are not able to detect events with the same level of sensitivity in a broader patient set. The results from these papers should therefore be interpreted with caution.

Three: Suitable patient exclusion criteria

The extent of cohort exclusion criteria can provide a measure of bias. The more exclusion criteria, the more potential bias (selection bias, see section 2.6). This is as it

may be easier to detect events with more exclusions (under more tightly controlled circumstances) but reduces how applicable a model can be in a real clinical setting. The majority of methodologies were based upon tightly restrictive eligibility criteria. Only five papers used non-restrictive exclusion criteria.

I observed that exclusion criteria could be grouped into two categories, 1) Clinical-based exclusions and 2) Data-based exclusions. Clinical-based exclusions were when patients were excluded due to some clinical detail. For example, if the patient did not have radical treatment (47). I believe that these exclusions were often to remove patients who may be more challenging to identify in the routine data. I believe these exclusions to be the most limiting, as excluding these patients could bias the results by driving up the sensitivity of event detection.

Data-based exclusions included removing those that had record linkage errors and those where gold standard data were not available to validate the routine data events. Other data-based exclusions included those who were not identifiable in the routine data (for example, not being enrolled in a healthcare plan) and others who had missing routine data (commonly found in administrative sources). These data-based exclusions also bias results because patients are being excluded from the analyses that are more difficult to identify in the routine data. This could increase detection sensitivity, by excluding the more challenging cases. Both types of exclusion reduce the feasibility of utilising such methods in a clinical trial setting – in a real clinical setting, patients cannot be excluded from follow-up if they do not fit desired criteria.

From the review literature, examples of excluded patients included; those not receiving all care at one particular hospital (Earle (163) and patients who had a recurrence within a year of diagnosis (McClish (164)). Further exclusions included, those that had not been enrolled for one year before and after diagnosis (unless death noted), to avoid identification of pre-existing comorbidities and initial cancer treatment and in addition, patients also had to be free of recurrence for 120 days after surgery to be eligible (Chubak (167)).

Patients were also excluded where claims data (administrative data, for example insurance claims) may have been incomplete (Kimmick (168)) and those diagnosed with

multiple tumour types or those where the cancer stage changed within 60 days after the index date (Nordstrom (169)). The patients excluded by Nordstrom may have more complex disease pathways and therefore may be more difficult to identify in the routine data. By not including these in the training set, the model is less likely to be able to identify more complex patients in practice. Patients were also excluded due to experiencing events within 180 days of diagnosis that could '*complicate identification of a recurrence*' (Lash (171)). Hence, this paper excluded the complicated pathways. However, in practice, in a clinical trial setting, patients cannot just be excluded from follow-up due to complicated data. Further exclusions included if the patient had not had cancer-directed surgery (Warren (176)), or if patients did not have radical treatment (Liu (47)). Again, limiting the generalisability of these algorithms in a wider patient cohort.

To summarise, the results of these studies therefore need to be interpreted with caution. When these algorithms are exposed to a non-restricted sample (for example, use in a clinical trial setting), the sensitivity of event detection may be impacted. These algorithms may not be fit for purpose in a clinical trial setting; any excluded patients would be permanently lost to follow-up.

In contrast to this, Lamont (165), Anaya (166), Ehrenstein (173), Haque (175) and Gupta (181) did not bias their patient sample with exclusion criteria. Of interest, Ehrenstein (173) excluded patients whose registry records contained distant metastases codes, however this was not to aid event detection, but was to identify events in the absence of this direct coding.

These five papers were designed to capture events for as many patients as possible. These algorithms are therefore more likely to be fit for purpose within a clinical trial setting. It is possible that the lack of exclusion criteria in these five studies led to low event detection accuracy (4/5 studies had an accuracy of less than 80%). If these algorithms were applied within a restricted cohort (as seen in the other papers), I hypothesise that the detection accuracy would increase, due to the detection of additional events. I believe these criteria to account for one of the greatest sources of weakness and bias.

Four: Tabular patient characteristics

Patient characteristics are important to confirm that results are generalisable to the general population. At the very minimum ClinicalTrials.gov require age and gender (191) to be defined when reporting study results. Therefore, age and sex/gender (unless specified in the text, for example, men with prostate cancer, women with ovarian cancer) were used as the indicators for the minimum that were required to be documented in a tabular form. The majority of papers displayed such a representation, however, some papers such as the study by Earle (163) did not. Lash (171) did not report sex and was therefore 'classified as negative' in the content analysis. Ritzwoller (182) also did not explicitly say the cohort was women with breast cancer. Perhaps the author felt that this was clear as they were reporting a breast cancer cohort. However, this is not necessarily the case as breast cancer in men is possible (yet rare).

Results with minimal patient characteristics should be interpreted with caution. The description of the cohort is vital to understand what cohort the algorithm could be applied to and to understand limitations of the study. For example, prior to utilising one of these algorithms in a clinical trial, the trial inclusion criteria would need to be compared to the sample in which the algorithm was developed. This would then help establish if the algorithm was capable of being used in the desired population. Therefore, the generalisability of the algorithms that did not publish patient characteristics is questionable.

Five: Patients in a randomised controlled trial (RCT)

Two pieces of literature included patients in an RCT (2/27, 7%) (figure 23) (Lash (171), Lamont (165)). The initial search screened to identify these RCT papers alone, however, this search was found to be too restrictive as these two papers were found. Hence, studies involving an RCT were credited in the overall risk of bias assessment, but the others were not excluded for completeness.

Lamont analysed patients in the CALGB 9344, doxorubicin dose escalation, with or without taxol, for node positive breast cancer trial. The trial data were linked to the routine Medicare claims data to measure disease-free survival (DFS). The DFS was

compared with the CALGB values to measure the algorithm accuracy (165). In the other study, Lash validated their algorithm in patients in the COLOFOL RCT, which compared two different regimes of follow-up after colorectal cancer resection (171). Lash plotted the cumulative incidence of events over time, as compared to the COLOFOL cohort. Neither studies used the administrative data to perform trial analyses, for example, replicating trial intervention efficacy analyses.

If the algorithms have not been tested in the clinical trial setting, the feasibility of using the algorithm has not been validated for clinical trial use. Hence, these algorithms would need to be re-validated using clinical trial data prior to implementation, as they may behave differently. For example, alternative treatment pathways may be being followed within the trial and the algorithm may not have been designed to capture these events. The practical and regulatory impact, such as linkage to trial number and patient consent, were also not possible to investigate during these studies. These details are important when evaluating the use of routine data within clinical trials.

Six: Multiple centre studies

Whether studies involved one or more centres was investigated. If data are analysed from a single centre, there is a risk of bias, due to coding and clinical patterns specific to institutions. In contrast, the use of multiple centres reduces the risk of bias. The impact of this is discussed throughout the thesis, for example in section 2.6.

17/27 (63%) (figure 23) of publications used multiple site data. It is possible that single-site studies (10/27) reveal higher detection sensitivities as differences in coding and clinician preferences may not be present in the small data sample. If these single centre studies were to be validated in another centre to the one in which the model was developed, there is a chance the sensitivity of detection could reduce. Clinical trials are often multi-site studies, hence, developing the algorithms using multi-site data is superior.

4.7.3.2 Intervention weakness and bias

Seven: Events compared to a reference

In order to validate the accuracy of the algorithm to detect events, the algorithm outcomes must be compared to a reference. All but one paper (176) compared the results of the algorithm to a reference (26/27, 96%) (figure 23). Warren (176) utilised no reference, therefore it is not known if the algorithm was able to accurately identify events.

If an algorithm has not been validated, it is not feasible for use; the algorithm by Warren needs to be validated prior to any further use. Without validation it is not possible to assess if the events being detected are true positives. Due to the importance of validation, the majority of studies performed this. In addition to this, the quality of the reference data also needs to be considered, this is discussed below (criterion fourteen).

Eight and nine

- *Eight: The algorithm was trained using a sample*
- *Nine: Validation was undertaken using an unseen dataset or statistically by cross-validation*

Statistical training of the algorithms were considered. *A priori* knowledge alone can be used without statistical training, but both can be used in combination. Training is undertaken by introducing code to data with the overall aim to identify particular patterns, to try to enhance model performance. There are various ways to train data, including statistically, for example, using logistic regression to identify variables predicting outcome but *a priori* knowledge can also be used, for example, choosing variables assumed to be important in identifying clinical outcome, without statistical confirmation.

18 of the 27 papers (67%) (figure 23) of the papers trained an algorithm statistically (plus or minus *a priori* knowledge). Training was used to identify variables of importance and to optimise the detection of events (to identify when variables should be identified). For

example, refining time windows for detection, such as the optimum treatment-free period needed prior to identification of recurrence from primary treatment.

To identify variables, methods in the literature included, logistic regression (McClish (164), Anaya (166), Hassett (180), Ritzwoller (182)), classification and regression tree analysis (CART) (Chubak (167), Nordstrom (169), Xu (186), Xu (187)) and random forests (Nordstrom (178)). To optimise the identification of events, studies investigated optimum backdating time windows to find the true date of event from the date of the proxy event (Ricketts (172)). Optimum time windows for detection were identified (Ricketts (172), Ehrenstein (173)) and peaks in code counts were investigated to enhance the timing of events (Hassett (180), Ritzwoller (182), Uno (185)).

The use of statistical methods improves the confidence in the algorithm performance. *A priori* knowledge alone may lead to missing events due to the routine data obscurities. For example, coding schemes and missing data may have implications on event detection if particular events are not documented; using *a priori* knowledge alone may not identify these obscurities. Using both statistical and *a priori* knowledge is the superior technique; clinical knowledge is important to ensure that the algorithm is not just identifying events due to the details in the coding. This combination technique was used by the majority of the studies.

Criterion nine assessed if validation was undertaken using an unseen dataset or statistically. The test set allows development of the algorithm, whilst the validation set enables a non-biased analysis and tests for overfitting of these data. These criteria were aiming to identify the studies that had undertaken validation on an unseen patient sample or using cross-validation techniques on one sample. If *a priori* knowledge was used alone to train the algorithm, then the initial sample was defined as suitable for validation, as these data had not been used to train the algorithm and therefore, overfitting was not a risk; one such example of this was the study by Lamont (165).

15/27 (56%) (figure 23) papers were validated using a separate sample of data (Xu (186) (187)) or cross-validation techniques. McClish (164) utilised a small subsample of

patients for validation, suggesting that a sample of the training data was utilised to validate the algorithm, potentially being at risk of overfitting.

It is vital to identify if the algorithm is capable of identifying events in an unseen population, prior to use within a clinical trial. An alternative cohort may contain different clinical or coding intricacies and therefore the algorithm may miss events if challenged with this new dataset. For example, the study by McClish (164) should be interpreted with caution, if the same data has been used for development and validation.

Ten: Algorithm can detect outcomes without pathology/lab reports or direct coding alone

The ability of an algorithm to detect events using routinely collected variables is of utmost importance to this project. Therefore, whether outcomes could be collected in the absence of detailed clinical data or directly coded diagnosis codes were assessed. 20/27 (74%) (figure 23) of papers were able to detect outcomes without these clinical data. The majority of papers using *direct* coding-only were excluded during the systematic review process. However, one paper linked clinical trial data to routine data and used *direct* coding alone (Lamont (165)) and was therefore included due to the importance of the clinical trial linkage.

Some studies had access to a pathology registry (Lash (171), Rasmussen (184)) and Ehrenstein (173) had access to lab reports for Prostate specific antigen (PSA), to help identify outcomes. All four papers from Denmark utilised additional clinical variables; no other countries utilised such variables. The Nordic countries are acknowledged for their population-based registries, which has enabled linkage of multiple longitudinal data sources across whole nations, hence the utilisation of such clinical variables in Denmark (192).

The use of additional reports reduces the generalisability of the algorithm. For example, the algorithms developed in Denmark cannot be used in the United Kingdom if the databases are not international. Hence, these algorithms could not be used in an international clinical trial setting. Where studies used international coding schemes, the

algorithms could be used internationally. One additional paper of interest (Haque (175)) only identified events that could not be identified from pathology reports.

Eleven: Investigation into the activity of variables

This analysis was to assess if the activity of variables was investigated, for example, frequency or clustering of events, as opposed to just presence of codes. 8/27 (30%) (figure 23) of the studies undertook activity analysis.

Focussing on the presence of codes only can lead to missing events, due to the nature of routine data. Some routine data events may be excluded intentionally but there is also erroneously missing and inaccurate data. Investigation into the activity of these data may therefore enhance the detection of events. This is discussed and the papers critiqued, below, through a thematic analysis (section 4.7.4). This is also investigated further within chapter 4.

4.7.3.3 Outcome weakness and bias

Twelve: The methodology developed allowed the timing of outcomes to be identified

Some papers purely sought to identify if a patient had experienced an event at any time and others sought to identify exactly when an event occurred. This was extremely important as the timing of event is vital for clinical trial analysis. 15/27 (56%) (figure 23) papers created an algorithm capable of identifying the time of event. A shorter delay in detection is sought after, especially if the algorithmically identified date is to be used for trial analyses.

Earle (163) detected relapse a median of 6.5 days after the medical record; Lamont (165) identified the date of recurrence for disease-free survival analysis with a median difference to the trial of 40 days; Liu (47) created survival curves to compare the algorithmically generated date to the gold standard and 16/20 of these were seen to have acceptable agreement; Ricketts (172) created survival curves to compare the timings to the reference, including 4/21 to the correct date; Lash (171) plotted cumulative incidence of recurrence against clinical trial data, and the incidence was seen to 'overlap significantly'; Ehrenstein (173) identified if the outcome was detected within 30 days from the true date and if it occurred at any time; Haque (175) identified the

date of subsequent breast cancer and for over 90% of patients the date differed to the reference by less than 60 days; Warren (176) identified the first indicator of recurrence and identified the median time to recurrence but no validation was undertaken; Hassett (180) created separate algorithms and identified where the code count peaked for 12 categories and integrated the values to identify a single time of recurrence; Gupta (181) found a median difference of 27 days; Ritzwoller (182) correctly classified 14.3% and 64.7% of events at three and six months, respectively; Wong (183) identified 4/16 within two weeks, 6/16 between four weeks and 100 days and 2/16 over 100 days; Rasmussen (184) found substantial concordance with the reference, 64% of events were estimated within 30 days; Uno (185) identified 75.3% within six months of the reference and Rasmussen (188) again found substantial concordance with a median difference of 17 days.

This highlights the non-standardised method of reporting the results, and the variable delays in detection. For example, Ritzwoller (182) only identified 14.3% of events within 3-months of the reference, compared to Earle (163) detecting relapse at only a median of 6.5 days after the reference. This demonstrates the difficulty in identifying oncology events from routinely collected data.

Where the timing of outcome is determined, it is possible to calculate surrogate endpoints for overall survival, for example PFS, for clinical trial analyses. In the two papers linked to RCTs (Lamont (165), Lash (171)), the timing of the outcome was assessed. Lamont identified the DFS and compared the individual events to the trial collected data to calculate the sensitivity at 2-years, 5-years and over the whole period of follow-up. Lash (171) plotted the cumulative incidence of recurrence against the clinical trial acquired data.

Twelve studies did not develop an algorithm that could identify the timing of the event. I would expect the sensitivity of event detection to be higher if the algorithm is only required to identify if a patient has had an event, or not, over the whole disease trajectory. Despite this, this was not a clear trend in the studies identified. This is potentially due to the chosen timelines to accept an event. There needs to be a

standardised definition to report the delays in detection, so that the summary statistics can be more comparable.

Thirteen: Algorithm feasible for use with regards to accuracy

The accuracy was investigated for the indirect algorithms in each study and the sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were compared, where specified. Warren and Yabroff (189) judged algorithms against an 80% detection accuracy rate for the statistics (sensitivity, specificity, positive predictive value), hence 80% being used as a proxy marker for quality here.

Using this definition for the published values, as opposed to the interpretation by the author, only 9/27 (33%) (figure 23) met these criteria. The larger studies tended to have lower accuracy, for example, Nordstrom (169), Hassett (180) and Ritzwoller (182), compared to the smaller cohorts such as, Lamont (165) and Liu (47). The single centres may have more complete and accurate coding, compared to the multiple site studies, possibly reflecting the difficulty in identifying events with larger samples (189). This highlights that, where possible, the analyses should be undertaken using multi-site data (as discussed above in criterion six), to confirm the integrity of the algorithms.

Lower accuracy algorithms should be interpreted with caution, as events could be missed, or more false positives could be present. However, the accuracy measure of greatest importance (for example, sensitivity, specificity), depends on the desired algorithm use. For example, if the routine data is being used alone, then a high sensitivity is vital to detect the maximum number of events. However, if a reference data set is available in addition, then the sensitivity may not be quite as vital. Here the positive predictive value be more important to reduce the number of false positives being identified.

4.7.3.4 General quality

Fourteen: The reference data are of high quality

In order to validate the accuracy of the algorithm-derived events, a quality reference is required, for example, note review, trial data and questionnaires. A 100% accurate 'gold

standard' is not possible (147), but a more accurate reference than the routine data is vital for validation. I deemed registry data to not be a suitable source of reference data to validate events; this is due to the derivation of registry data as discussed in section 1.2.3. Registries are often utilised as the routine data source to be validated and as such should not be used as the reference.

23/27 (85%) (figure 23) of papers used a quality reference source. Those without a suitable reference included McClish (164), who used a registry plus a small record review; Warren (176) who used no reference and Rasmussen (184) who used a Danish bladder cancer registry in the first paper and in the second paper (188) again used a registry plus GP and small medical record review, where no event was identified in the registry.

The accuracy statistics for these four studies should be interpreted with caution. If the reference itself has missed events, then events detected by the algorithm may actually be true positives and not false positives. Likewise, if the reference has false positive events, events missed by the algorithm may appear to be false negatives rather than true negatives. Inaccurate measures of algorithm performance may be generated with a low-quality reference; this leads to uncertainty of the algorithm performance. However, as mentioned above, there is no 'true' reference available (147) but the higher the quality of the reference, the greater the algorithm integrity.

Fifteen: Table/figure comparing algorithm results and the reference data

A table or figure outlining the accuracy of the event detection algorithm is important to allow these data to be more easily interpreted. Despite, this only 17/27 (63%) (figure 23) displayed these data in this format, for example, in a figure (Chubak (167)) or in a table (Lamont (165), Gupta (181)).

Where algorithms are being assessed for clinical use, accessibility and transparency of results is vital. If reporting guidelines requested comparable items, then it would be easier to make assessments of algorithm performance across studies.

Sixteen: Indexed by PubMed

27/27 (100%) (figure 23) of the articles were indexed by PubMed. This was utilised as a proxy, although arbitrary, marker for quality of the journal of publication. This was used as an alternative to the journal impact factor, as this marker of quality has been previously questioned (193).

No comparisons could be made for the quality of the papers here as all were indexed by PubMed. However, this suggests that all of these papers were published by journals that met '*vigorous review or selection criteria*' (194). This, therefore, gives a proxy marker that the algorithms have been reviewed by an independent researcher. This suggests that the algorithm designs have been reviewed and therefore these algorithms may have higher integrity.

Seventeen: Algorithm variables published (or summary)

In order to replicate analyses, the variables used in each algorithm should be published. Despite this, only 19/27 (70%) (figure 23) published the variables.

If the studies do not report the variables, it may not be possible to replicate the algorithm and validate or use it within your cohort. The algorithm is not generalisable if it is not possible to replicate. It may be possible to request the algorithms off the authors, but this is often also not clear. It may not be possible, or appropriate, to publish the entire algorithm but here it should be obvious that it is possible to request the code or variables off the author. This enables the algorithms to be replicated within similar, or different, cohorts for further validation and use.

4.7.3.5 Overall weakness and bias assessment

In assessing these criteria through analysing participant, intervention, outcome and quality factors, it was found there was a varying level of weakness and bias. Greater than 85% of studies fulfilled the following criteria: the cohort was accurately identified (criterion one); the events were compared to a reference (criterion seven); the reference data were of high quality (criterion fourteen) and the studies were indexed by PubMed (criterion sixteen) (figure 23).

In contrast, alarmingly less than 20% of studies used suitable patient exclusion criteria (using my defined criteria) or linked the routine data to an RCT. The patient exclusion criterion was deemed to therefore be the greatest source of weakness and bias, as the accuracy of event detection was hypothesised to be increased, due to excluding more challenging cases (figure 23). Within a real clinical setting, patients cannot be excluded from follow-up due to less *favourable* disease trajectories. The lack of RCT routine data-linked studies was also a major finding (figure 23).

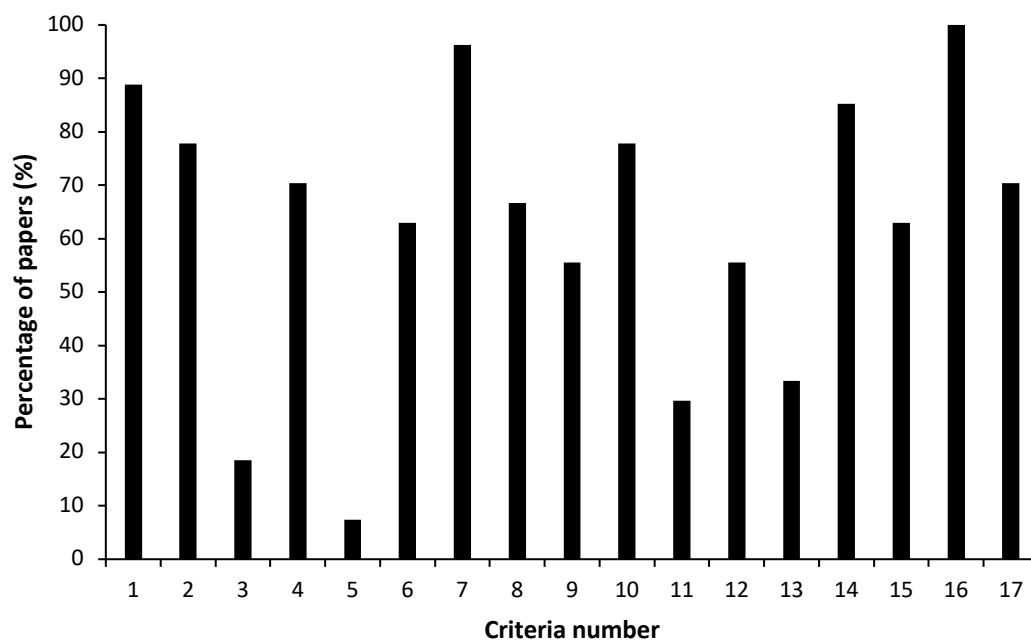


Figure 23: The percentage of papers fulfilling each of the 17 criteria analysed during the content analyses.

At a study level, the papers identified to score the lowest level of weakness and bias were by Haque (175) and Chubak (167), scoring 14/17, 82.4% of criteria. The papers identified to have the highest level of weakness and bias were by Hagberg (170), Deshpande (174) and Livaudais (177) (figure 24).

The studies by Haque (175) and Chubak (167) also identified a maximum sensitivity of 0.97 and 0.99 of events and hence, were critiqued to be the highest performing and

most feasible algorithms (both studies were in breast cancer using databases in the USA).

The study by Livaudais (177) showed the highest sensitivity of event detection (1.0) but also scored the lowest in the weakness and bias assessment. This is potentially due to the bias in the studies, enhancing event detection.

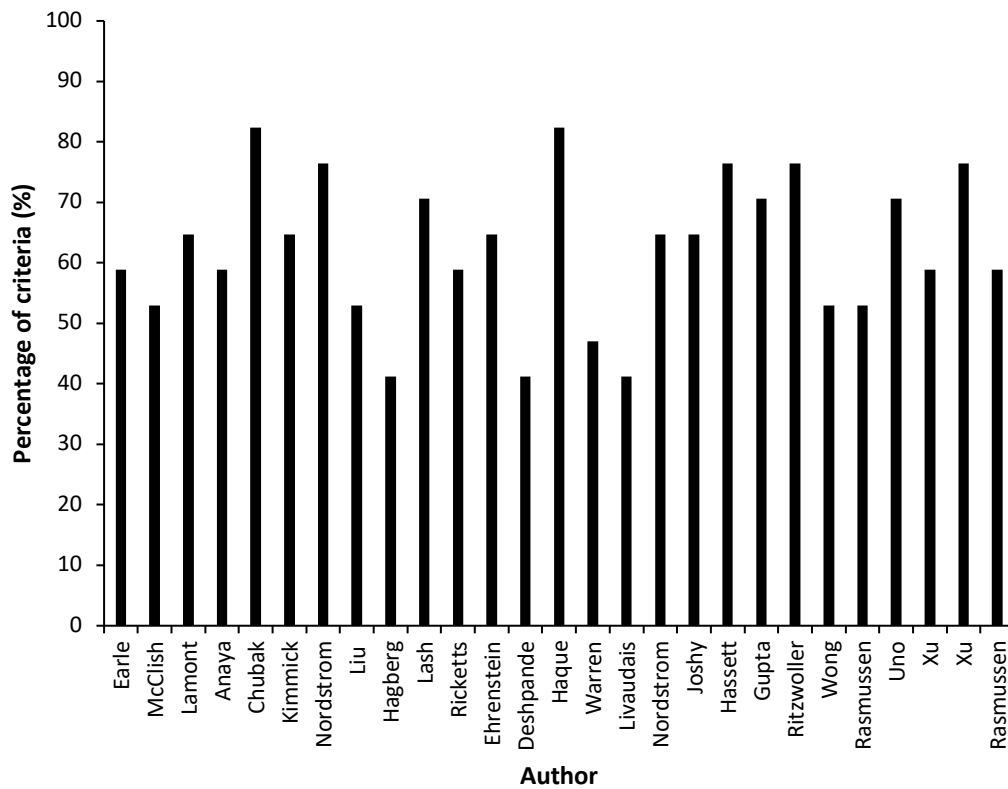


Figure 24: The percentage of the 17 criteria fulfilled by each author.

Where 0% and 100% = 0/27 and 27/27 of the criteria being scored, respectively. The total number of criteria scored, by author, are in the appendix section 8.2.2.

4.7.4 Methodology analysis summary

Further analysis was undertaken into the development of the algorithms for: 1) the type of variable utilised in the algorithm (outcome coding, procedure coding, clinical coding and activity coding) and 2) the algorithm design.

4.7.4.1 Types of variable

A thematic summary identified four different methodological technique themes to identify events:

Type of variable	Examples
Outcome*	Secondary codes
	Recurrence
	Metastases
	Neoplasm codes
	Cause of death or date of death
	End-of-life care indicator (hospice)
	Pathology codes
Procedure	Surgical
	Non-surgical (chemotherapy, RT, imaging, biopsy)
Clinical*	PSA
Activity	Number increase (visits, imaging)

Table 44: The variable coding themes identified from the literature; outcome, procedure, clinical and activity.

**: Algorithms developed to use clinical or outcome variables alone would be excluded, unless the algorithm was developed linked to RCT data (see the exclusion criteria in table 41).*

The most common design (table 44), was utilising both outcome and procedure coding (13/27, 48%) (table 45). The procedures were used as proxies to enhance the detection of events that would have been missed using outcome coding alone. The second most common technique was using procedure codes alone (7/27, 26%) without the outcome coding.

Lamont (165) was the only paper to utilise outcome coding alone, due to the exclusion criteria (table 41). The study by Ehrenstein (173) was the only study to utilise clinical diagnostic test values as well as procedure codes. 5/27, 19% of papers investigated the activity of variables and not just the presence of the variables, however, these papers

still relied upon outcome coding. Not one algorithm using activity markers was designed without the use of outcome coding (table 45).

Algorithms developed with outcome coding (especially alone) should be interpreted with caution. Many routine clinical datasets are collected for primary reasons that often do not include research. Hence, diagnoses coding is often not mandatory; this can lead to sporadic reporting. Therefore, these techniques may reduce the algorithm performance in detecting events. In the HES data, procedure codes are documented with a date (for example, date of the operation or the admission) but diagnoses codes are not reported with a date. Hence, if the diagnoses code was being used as an indicator of the event date, this could be delayed from the true diagnoses date.

Author	Outcome alone	Procedure alone	Outcome & procedure	Clinical & procedure	Outcome & Procedure & Activity
Earle (163)		+			
McCLish (164)			+		
Lamont (165)	+				
Anaya (166)			+		
Chubak (167)					+
Kimmick (168)		+			
Nordstrom (169)			+		
Liu (47)		+			
Hagberg (170)			+		
Lash (171)			+		
Ricketts (172)		+			
Ehrenstein (173)				+	
Deshpande (174)			+		
Haque (175)					+
Warren (176)			+		
Livaudais (177)		+			
Nordstrom (178)			+		
Joshy (179)		+			
Hassett (180)					+
Gupta (181)		+			
Ritzwoller (182)					+
Wong (183)			+		
Rasmussen (184)			+		
Uno (185)					+
Xu (186)			+		
Xu (187)			+		
Rasmussen (188)			+		
Total	1	7	13	1	5

Table 45: The methods utilised for algorithm development by the author.

If one algorithm in a study was developed using procedure coding alone, the study was classified as so (despite developing other algorithms, for example, plus outcome).

The five instances where activity of the variables were investigated were, Chubak (167), Haque (175), Hassett (180), Ritzwoller (182) and Uno (185). Chubak (167) and Haque (175) used the frequency of a single event (visits) to fulfill a categorical variable. For example, if four visits were met within a defined time period, the recurrence indicator was identified, flagging an event. The final algorithms contained a rule to flag outcomes that identified the presence of two visits, plus a code for a secondary malignant neoplasm within 60 days. Although not included in the final algorithm, the number of instances of breast imaging and diagnoses codes in certain time periods, for example, two visits for mammography within 60 days were also investigated. The final Haque (175) algorithm included identifying the number of oncology visits in 60 or 90 days (four visits within 90 days and three visits within 60 days) to flag an outcome of interest.

In contrast, Hassett 2017 (180), Ritwoller 2017 (182) and Uno 2018 (185) (all papers by the same team) investigated the ability of using unique peaks in the code count to detect events. Hassett (180) initially developed an algorithm based upon twelve potential indicators of cancer recurrence to identify outcomes of interest and to identify the timing of the events.

Indicators of recurrence were assessed to include the mean numbers of events, for example, imaging and inpatient events per year. Indicators were usually categorical variables, unless the absolute code count increased over time, in which these were classified as continuous indicators. To determine the timing of the event, the period when the code count peaked for each individual variable was assessed. Therefore, it was possible for a patient to have twelve different recurrence times. Subsequently a single time was statistically derived from the twelve dates to flag the outcome of interest.

Ritzwoller (182) in the following year enhanced the Hassett (180) algorithm. Algorithms were separately developed to 1) identify events and 2) identify the timing of events. To identify events, the number of secondary malignant neoplasm codes were determined; if the absolute number reached the threshold (34 events), the probability of cancer recurrence was 100%. In those patients where 34 events were not identified and the threshold was not reached, indicator variables were utilised (as in the Hassett (180) study), to generate a probability of having recurrence. The variables to indicate

recurrence included a mean increase in the imaging code count over the year and the total number of chemotherapy codes, radiotherapy codes, hospice codes and mastectomy codes.

The timing of recurrence was based upon the Hassett (180) study, identifying where the code count peaked for each individual code group. Each variable held a weight, and this allowed a final estimate to be identified. An offset (difference between the time when the code count peaked and the gold-standard) was investigated in addition, to aim to identify a recurrence date nearer the 'true' value.

The final paper by the team (Uno, (185)) also aimed to identify the timing of event. In addition to enhancing the previous timing model, a method was developed to infer the date of missing events. This was done by using the time halfway between diagnosis and the end of follow-up, as the date of recurrence.

The use of activity variables may have enhanced algorithm performance; where sensitivity values were published (Chubak (167), Haque (175), Hassett (180), Ritzwoller (182)), the sensitivities ranged between 0.77 (Hassett (180)) and 0.99 (Chubak (167)). No studies investigated the cumulative sum of all cancer-related interactions, but instead looked at the individual variables in isolation. Due to the absence in the literature and routine data limitations, this cumulative method of identification was investigated further. The hypothesis was that events could be identified despite the absence of *direct* outcome coding in the routine data.

4.7.4.2 Algorithm design

As discussed above in section 4.7.4.1, different algorithms are designed to screen for different types of codes (the outcome, procedures and clinical variables). However, in addition, the algorithms are also designed in other ways. For example, the algorithms begin screening from different time points and often require treatment-free intervals. Figure 25-27 summarise the different algorithmic designs.

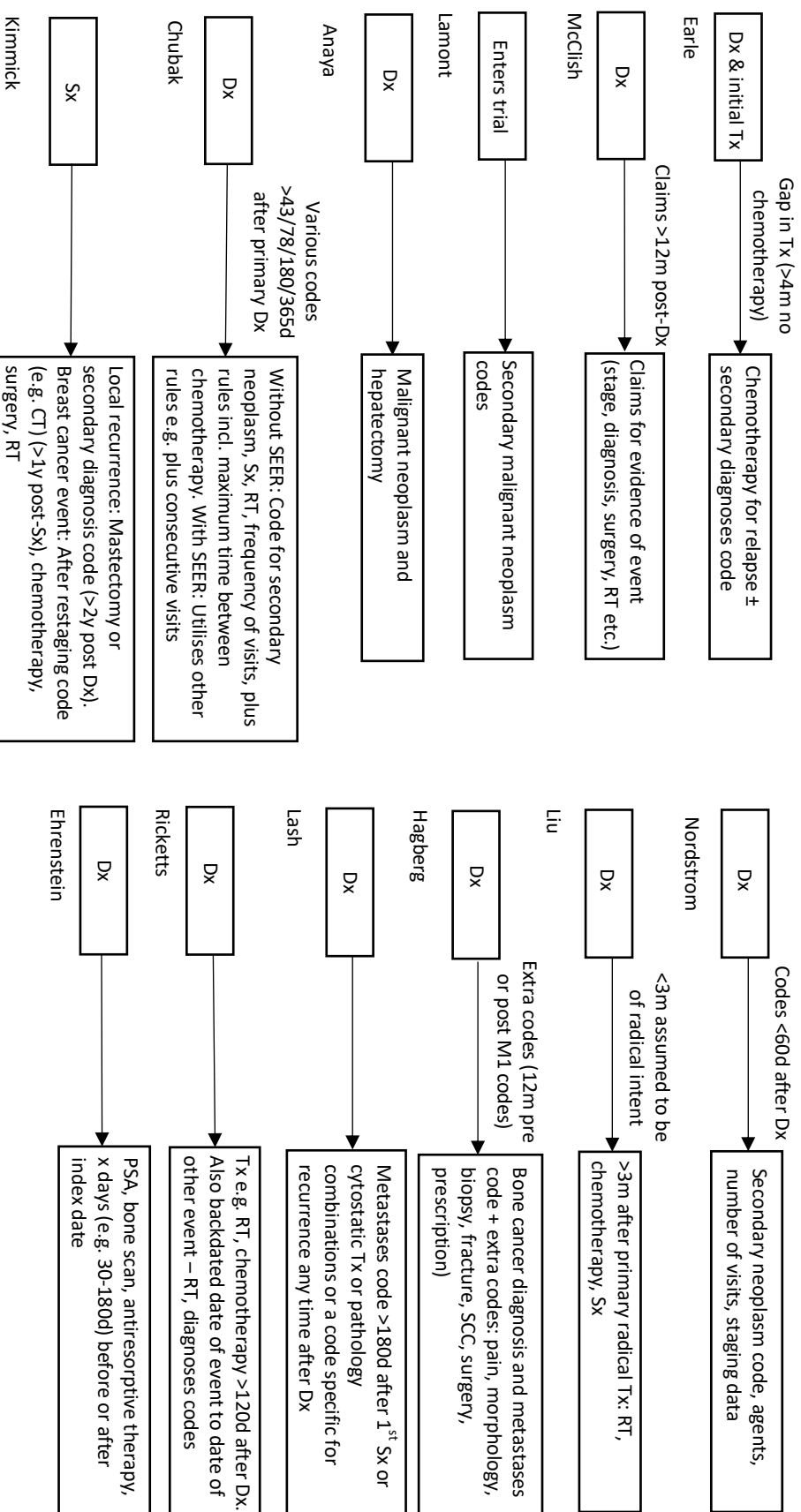


Figure 25: Part 1: Summary of the concepts identified during the full-text review of 12/27 papers analysed in the review; Sx = surgery; Dx = diagnosis; Tx = treatment. RT = radiotherapy; HTx = hormone therapy; M1 = metastatic; TNM = tumour, node, metastases. The references correspond to the order of those in table 45.

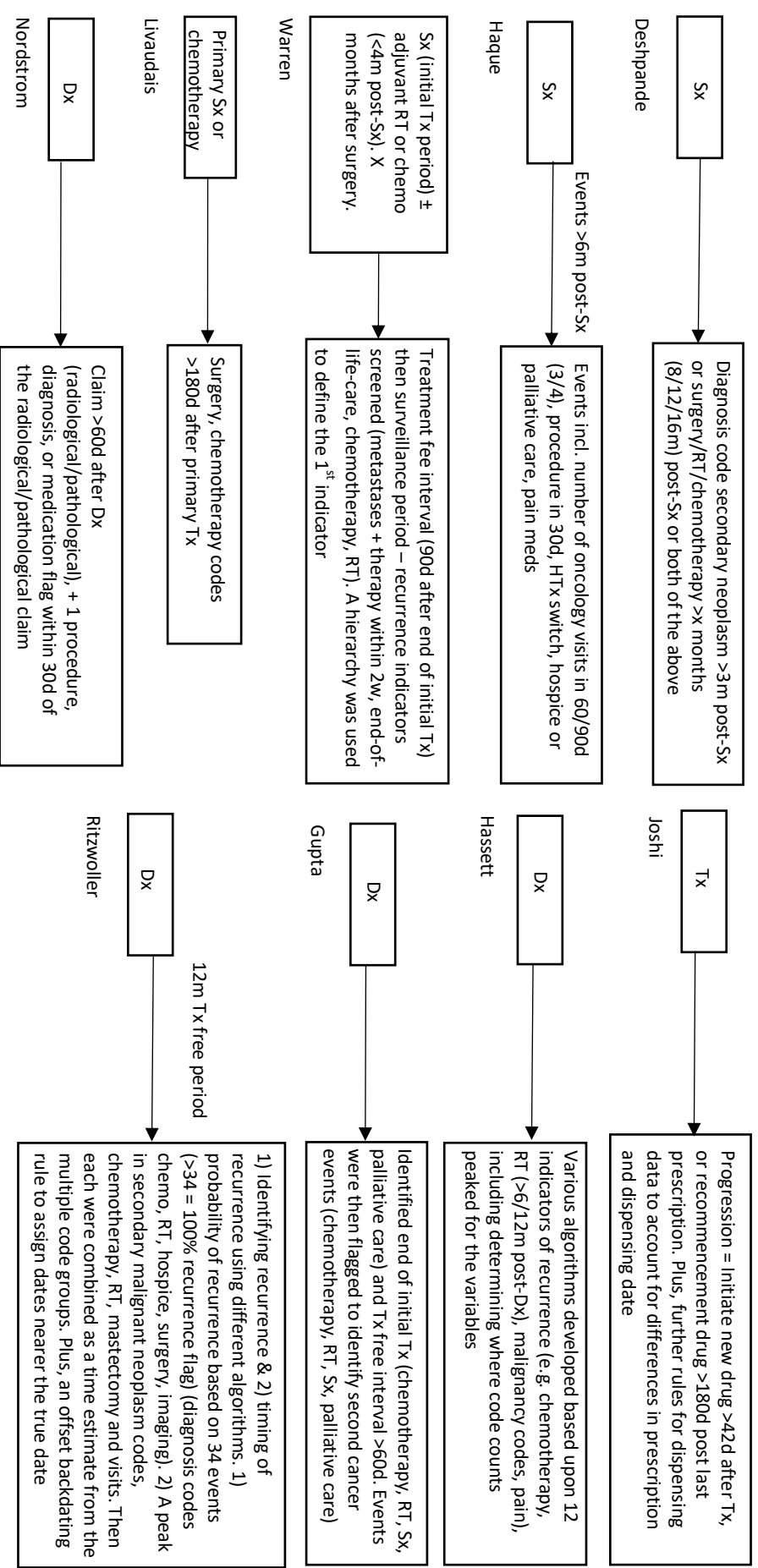


Figure 26: Part 2: Summary of the main concepts identified during the full-text review of 9/27 papers analysed in the review; Sx = surgery; Dx = diagnosis; Tx = treatment. RT = radiotherapy; HTx = hormone therapy; M1 = metastatic; TNM = tumour, node, metastases. The references correspond to the order of those in table 45.

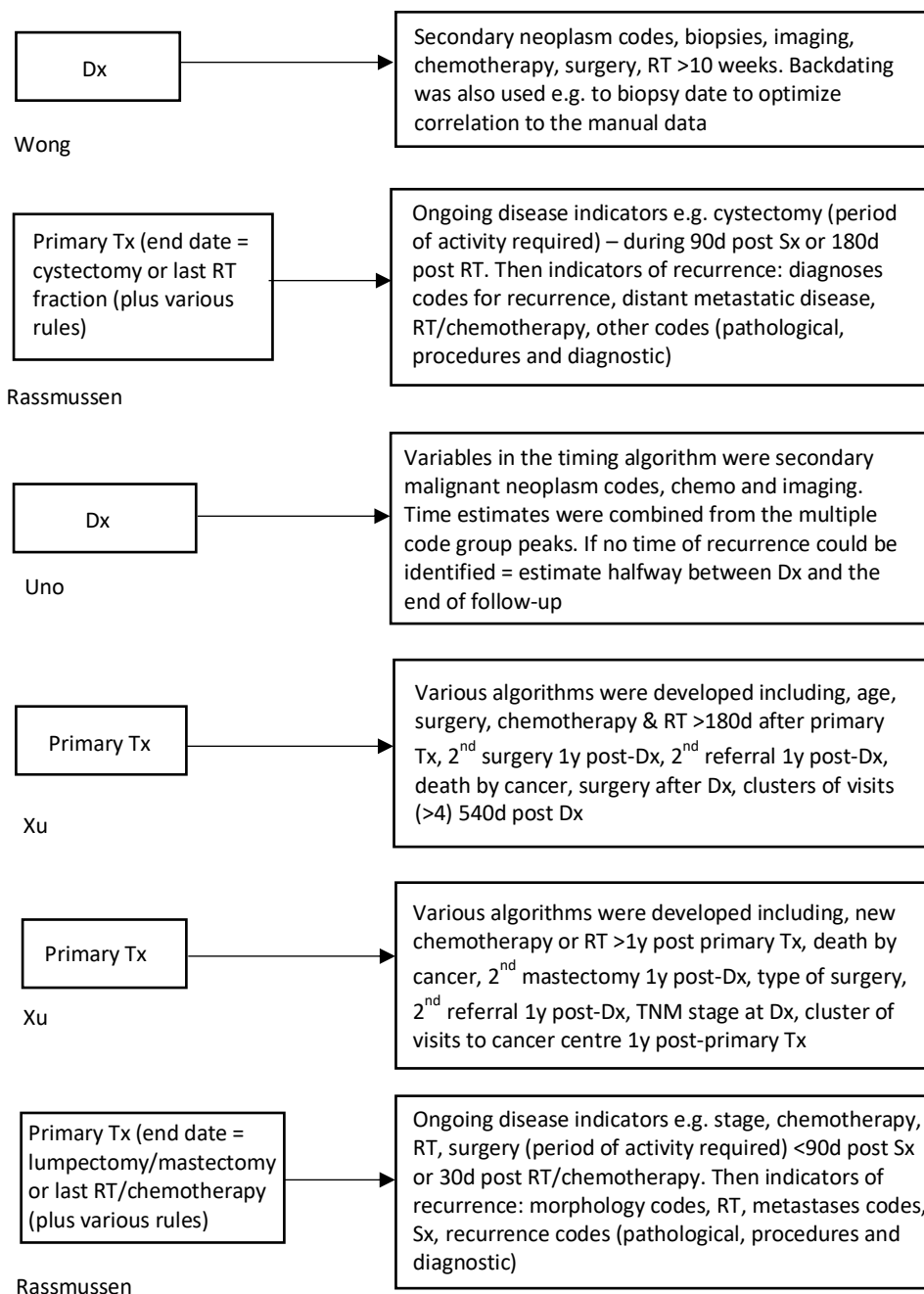


Figure 27: Summary of the main concepts identified during full-text review of the 6/27 papers analysed in the review.

Sx = surgery; Dx = diagnosis; Tx = treatment. RT = radiotherapy; HTx = hormone therapy; M1 = metastatic; TNM = tumour, node, metastases. The references correspond to the order of those in table 45.

Treatment free intervals were often required to fulfil the algorithm rules to detect an event (for example, Earle (163), McClish (164), Chubak (167), Haque (175), Ritzwoller (182)). The aim of this is to enhance the identification of events, however, it is possible that this may restrict event detection; events may occur within that time window. For example, a patient that progresses very quickly after diagnosis may be excluded. This is in contrast to one study by Nordstrom (169) where the codes to identify events were required to be present in the first 60 days after diagnosis. This means that events occurring sooner can be identified, but events occurring during long-term follow-up would be excluded.

Codes are also screened to identify events from different time points. For example, diagnosis (166), entering a trial (165), primary treatment (179) and specifically surgery (168) or chemotherapy (177). Where primary treatment is used as a marker to begin screening codes, this would exclude any patients that experienced an event prior to treatment. However, allowing screening to occur from diagnosis may increase the number of false positives (reduce the PPV) whilst enabling detection of earlier events. This may be illustrated by three low PPV performing studies (Ehrenstein (173), Nordstrom (178) and Chubak (167)) that screened from diagnosis. This can be compared to studies with high performing PPVs, such as by Rassmussen ((184), (188)) which screened from primary treatment (table 43, part 2). There is therefore a trade-off between developing an algorithm with a high sensitivity and a high PPV. Detecting events earlier at diagnoses may increase event detection (increase the sensitivity) but may lead to more false positives (reduce the PPV).

In addition, to optimise the identification of events, Ricketts investigated optimum backdating time windows to find the 'true' date of event from the date of the proxy event (172). It is possible that the new backdated event may also not be the 'true' date. For example, the radiotherapy date may not flag the true date of the cancer event.

Algorithms were designed differently to try to enhance event detection accuracy. However, what was shared across studies was that all codes were screened from an index event (treatment or disease event) throughout the disease trajectory, to identify outcomes.

4.8 Systematic literature review: Conclusion

This systematic review identified different methods to detect outcomes of interest, where the majority of outcomes were detected in non-trial patients. Worldwide, only two studies linked routine data to trial data, but the timing of the outcome was assessed in both of the papers. However, neither of these models were designed to detect outcomes without clinical databases containing details such as pathology or *direct* coding alone, limiting their generalisability. Only one study compared the cumulative incidence of algorithmic-detected outcomes to trial data collected outcomes but the ability to identify trial level surrogacy (treatment intervention efficacy, see section 4.9.5) was not assessed. In addition, this study only involved a small sample of 63 events in colorectal cancer. Hence, no studies worldwide have developed a clinically useable validated tool for trial use. It was also identified that no UK routine databases have been used to perform oncology clinical trial analyses.

Therefore, this review confirmed there remains an unmet need to utilise routine datasets alongside clinical trial data, to identify oncology outcomes that could be used to perform surrogate endpoint analyses.

In clinical trials, time to outcomes are a commonly collected trial measure and therefore a method to extract this information from routine data would uncover hidden but routinely available, cheap, rich data to allow further oncology questions to be answered.

4.9 Materials and methods

4.9.1 Approvals

All STAMPEDE data, routine HES data (Public Health England (PHE) and National Health Service Digital (NHSD) sourced) and clinical noting data were sought with relevant ethical and regulatory processes in place; see the section 2.4 for the detailed list of approvals.

4.9.2 Participants

Participants were included across four cohorts (table 46); some patients overlapped between groups, but the number of novel participants in each cohort can be seen in table 46. Cohort one to three (including the cumulative cohort) included eligible England STAMPEDE participants recruited from the University Hospitals Birmingham Queen Elizabeth Hospital (UHB QEH), selected from STAMPEDE treatment arms A-F (introduction, 1.5.1) (72). The patients in cohort one were chosen by M. Gannon at the MRC CTU at UCL as a sample of patients that were known to have routine data available and at least one skeletal related event (SRE) instance (to also enable SRE outcome availability analyses), four of which had spinal cord compression (an SRE). The single-site data were utilised due to only having access to single-site routine data initially. There was no resource or regulatory access to perform validation from outside of UHB QEH. Cohort two and three patients were chosen as a random sample of the remaining STAMPEDE patients linked to the routine data for training and validation; these patients were not restricted by SRE occurrences. The largest (cohort four, N=1,695) validation study was undertaken in an England multi-site cohort. This included patients on STAMPEDE treatment arms A, C and E (introduction, 1.5.1) (Table 47) (figure 28).

Cohort number	Stage	Total cohort size	Novel patients
1	Model development (pilot/feasibility)	6	6
2	Algorithm training	44	41
3	Algorithm validation 1	46	46
-	Algorithm validation 2 (cumulative) (cohort 2 & 3 plus 3 from cohort 1)	93	0
4	Algorithm validation 3	1,695	1,644

Table 46: The number of patients (total cohort size and number of unique/novel patients at the time of the analyses) in each cohort by the stage of analyses.

The number of patients in the entire STAMPEDE docetaxel comparisons can be seen and the number utilised in the largest cohort four analysis (figure 28).

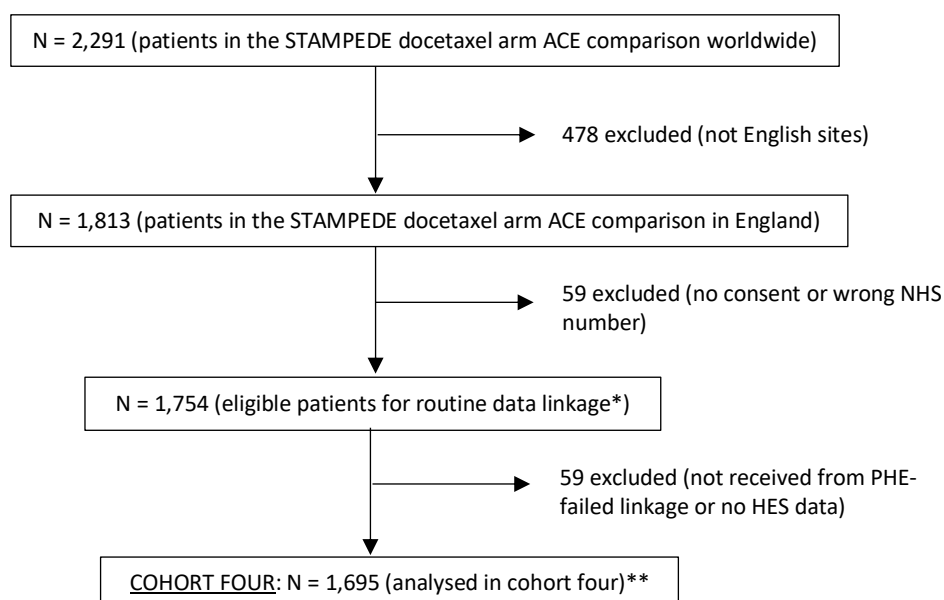


Figure 28: Diagram illustrating the final cohort selection (cohort four), from the total in the docetaxel comparison in the STAMPEDE trial.

*= This includes patients providing a name at randomisation, implicitly giving consent for flagging, who have not withdrawn consent or have a wrong NHS number. **= Those linked by PHE, with HES data available.

4.9.3 Data extraction, outcome measures and processing

Data were extracted from three separate sources: the STAMPEDE trial, the routine data and the clinical records. The records extracted from STAMPEDE included: date of randomisation, treatment allocation, trial visits and trial outcomes (for example, progression). These data for the six initial pilot cases were sourced directly from the

case report forms (CRFs) but for all other stages, data were extracted from the trial database. The required fields (table 47) to identify individual routine data interactions were linked to each unique STAMPEDE trial ID.

The STAMPEDE data were also used to extract patient characteristics, including; treatment arm, broad disease grouping, age and PSA at randomisation (interquartile range (IQR), median, range) and tumour (T), node (N) and metastases (M) category at randomisation. I calculated the summary statistics using code that I wrote to extract the information from the STAMPEDE database. C.Brawley at the MRC CTU at UCL, extracted the age summary statistics for the largest multi-site cohort, because I did not have access to the non-UHB QEH date of birth (DOB) values.

Variable	Code
Date the patient was admitted to hospital/date appointment scheduled	ADMIDATE/APPTDATE
Date the patient was discharged from hospital	DISDATE
The specialty under which the consultant is contracted	MAINSPEF
The type of interaction – admitted patient care or outpatient event	ADMITYPE
Diagnosis codes - ICD-10	DIAG_01 – DIAG_12
Operation codes - OPCS version 4.2 to 4.8	OPERTN_01 – OPERTN_12
The organisation acting as the healthcare provider	PROCEDURE

Table 47: The routine data (HES) fields analysed during model development, to identify prostate cancer-related events.

Case series of the clinical records were undertaken using the clinical noting data and this was considered the reference standard for all analyses, except for trial endpoint analyses where the STAMPEDE records were considered the reference standard.

Records of interest extracted from both the routine data and the clinical records upon note review included: date, type of healthcare interaction (inpatient, outpatient) and diagnostic, treatment and follow-up details. In addition to this, progression outcomes were identified from the clinical records and defined as, local or nodal progression, distant metastases and/or skeletal-related events (SRE). Dates for outcomes were assigned as the date of objective detection, unless not specified, whereby the dates of SRE detection or biochemical failure were assigned.

During model development, for each participant, each clinical record and routine data interaction was manually inspected and tabulated to classify as 1) prostate cancer-related record, or 2) non-prostate cancer-related record, which were removed.

4.9.4 Data censoring

Where the three data sources were being initially investigated individually (during model development – cohort one), each of the three data sources were analysed until the end of follow-up for each source. However, where comparisons were being made using multiple data sources, analyses of the data sources were undertaken from the date of randomisation into STAMPEDE until the last available mutual event, identified in either the routine data, the clinical records, or the STAMPEDE data, depending on what analysis was being undertaken.

Throughout endpoint analyses during algorithm training (cohort two) and the small-scale validation (cohort three), the STAMPEDE May-2015 dataset freeze for arms B, C, E and contemporaneously recruited arm A patients were used (data frozen 13/05/15). For the D, F and corresponding arm A patient analysis, the December-2015 data freeze was used (data frozen 15/12/15). This date was substituted for the last routine data or clinical record event if these events preceded the STAMPEDE censor date.

In the large-scale validation (cohort four), more recent HES data (up to the 31/01/18) was available compared to the STAMPEDE data; hence, if no earlier outcome was identified using the HES model, or the HES outcome that was found occurred after the STAMPEDE censor (STAMPEDE arm A, G data freeze December 2017), the event was censored back at the individual STAMPEDE censor date (clinical noting analyses was not possible in the multi-site cohort four analyses).

4.9.5 Requirements to develop a surrogate endpoint

In order to validate the HES-derived endpoint (event/outcome of interest), previously published criteria were followed (195, 196). This involved testing the endpoints at an individual level and at a treatment level (195, 196). These criteria can be seen in practice in the International Intermediate Clinical Endpoints in Cancer of the Prostate (ICECaP)

study (98). These criteria require a two-stage validation model investigating validating that a surrogate can show a strong relationship between:

1. Individual endpoint effects (patient level surrogacy) – association of the novel endpoint to standard endpoints, without reference to an intervention (98, 195, 196) (151). Within this study, the algorithm generated endpoints were compared to the standard STAMPEDE endpoints and correlation coefficient analyses were undertaken; further details can be seen below.
2. Endpoint treatment effects (trial/treatment level surrogacy) - a strong correlation between treatment effects on the novel endpoint was required (98, 151, 195, 196). Hazard ratio analyses were conducted; further details can be seen below.

4.9.6 Analytical methods

A summary of analytical methods can be seen in table 48, described in further detail below.

Analyses	Model development (N=6)	Algorithm training (N=44)	Algorithm validation (N=49, 93, 1,695)
Case series (interactions classified as prostate-cancer related or not)	✓	✓	✗
Records of interest extracted from the routine data and the STAMPEDE data (classified as prostate-cancer related or not)	✓	✓	✗
Routine data accuracy & SRE accuracy	✓	✗	✗
Clustering analyses (varying time intervals, varying thresholds, coding variations, removing early confounding events, removing overlapping codes, removing routine oncology appointments, assessment of continual progressive states)	✓	✓	✗
Surrogacy analyses: Patient level surrogacy (association of the algorithm endpoint with the STAMPEDE endpoints) ± magnitude of the difference analyses ± Pearson's correlation ± Bland-Altman analysis	✓	✓	✓
Surrogacy analyses: Trial level surrogacy (comparing hazard ratios – comparing treatment effects)	✗	✓	✓

Table 48: A summary of the methods used during the analyses.

Firstly, routine data accuracy calculations were performed for: overall accuracy of detecting prostate cancer-related events (missing event analyses); investigation into the

feasibility of an indirect model, by investigating the effect of inferring missing HES events (inferred event analyses) and the effect of intentionally excluded codes on the data quality (excluded event analyses) (table 49).

Accuracy analysis	Aim	Included events	Excluded events
Missing events	Validate the HES against note review-derived records to assess the data accuracy	All prostate cancer related events (inpatient and outpatient)	Unconfirmed events that could not be validated, including, STAMPEDE nurse visits (CRF identifiable but often not documented in clinical noting) Did not attend or cancelled appointments (event did not occur)
Inferred events	To assess if further events could be identified using alternative coding	Events included in the 'missing event' accuracy calculation, plus: Inferred coding for events: <ul style="list-style-type: none"> • MAINSPEF 800 (clinical oncology outpatient event) = radiotherapy • MAINSPEF 960 (allied health professional outpatient event) = scans 	Those excluded from the 'missing event' accuracy calculation
Excluded events	To assess whether the HES accuracy (missing events) was due to missed or planned exclusions	Events included in the 'missing event' accuracy calculation	Those excluded from the 'missing event' accuracy calculation, plus: <ul style="list-style-type: none"> • X-Ray imaging • Trial treatments

Table 49: The events included in the three accuracy calculations, for: missing, inferred and excluded events.

MAINSPEF = a HES field which documents the speciality which the patient's consultant is contracted to (table 47).

During the model development (pilot), we hypothesised that if a patient was experiencing a clinically relevant outcome (for example, progression), the number of healthcare events (for example, contact with services, such as scans and visits) would increase, providing us with a clinical footprint. The aim was to capture periods of activity above routine oncology (197) and STAMPEDE trial monitoring (198) which could identify trial outcomes of interest, such as progression. To build a model to identify these outcomes of interest, the counts of records were generated from both the clinical records (initially) (see table 50 for included and excluded events) and the routine HES tabular processed data (after the clinical record analyses) to represent graphical metrics

of activity. If the level of activity fulfilled the defined model rules (table 51), a routine data derived outcome was identified. The time to this HES-derived outcome was proposed as the ‘activity-free survival’, the AFS.

Included events
Medical oncology interactions (inpatient, outpatient)
Clinical oncology interactions (inpatient, outpatient)
Urology interactions (inpatient, outpatient)
Nurse led STAMPEDE clinic visits
Cancelled appointments
Did not attend appointments
Radiotherapy
Trial chemotherapy (for example, docetaxel, zoledronic acid)
Non-trial chemotherapy (for example, strontium, radium)
Imaging (including X-rays)
Admissions related to prostate cancer
Excluded events
Unplanned visits
Outpatient or inpatient non-prostate cancer-related events (e.g. cardiology)
Blood tests (e.g. prior to chemotherapy, PSA)
Specialities related to side effects e.g. maxillofacial appointments - osteonecrosis jaw

Table 50: The events that were identified during the pilot analyses and were included and excluded from the clinical noting data to build the HES-outcome detecting model.

N.B This is not an exhaustive list of potential events (compared to table 52) as not all events could be identified within the initial six prostate cancer patient trajectories.

Whilst triangulating the three data sources, model rules were refined for optimum outcome detection, to identify when a cluster of events was deemed to be a ‘routine data outcome’. This included varying time intervals to cluster the healthcare interactions, a threshold number of events to reach to identify the timing of the outcome (for first and subsequent outcomes) and coding variations to enhance the detection.

For each of the candidate thresholds and time intervals (seven different time intervals were investigated, using both clinical and statistical assumptions) calculations included, true and false positive outcomes, sensitivity, PPV and false negatives. In addition, the mean difference in detection (the mean magnitude of the difference) was also calculated between the clinical noting derived outcome and the routine data outcome. The accuracy of detecting SRE outcomes *directly* from the routine data coding was also investigated in all six pilot patients (analyses also included splitting the cohort into the three earliest and latest patients, to see if the accuracy was time-dependent), by

comparing the routine data-derived outcomes to the clinical noting and trial-derived outcomes.

Subsequently, during algorithm training (cohort two), the model was automated within R v3.3.2 (109) with RStudio v1.0.136 (110) and applied to the routine data (code available on request). Two different sets of analyses were undertaken to assess *patient level surrogacy* and the magnitude of the difference in detection (weeks) for all true positive STAMPEDE and routine data endpoints were calculated.

- 1) All reference clinical noting progression outcomes were compared to the routine data-detected outcomes and the STAMPEDE (PFS, MFS and FFS) documented endpoints (99) (the ability to detect the first two outcomes per patient, are presented). The clinical noting and STAMPEDE endpoints were also assigned to the algorithm time intervals to enable a direct comparison.
- 2) The *first* routine data outcome per patient (the AFS) was compared to the STAMPEDE endpoints (MFS, FFS, PFS) (the reasons for the missed or delayed routine data outcomes were identified).

The ability of the algorithm to correctly classify the overall outcome status over the whole of follow-up was also assessed (for example, identifying if the patient did or did not experience an outcome during the whole of the follow-up period). In addition, the total number of routine data-derived outcomes (AFS, SRE) was compared to the number of standard STAMPEDE detected outcomes (FFS, MFS, PFS, SRE). Subsequently *trial level surrogacy* was investigated by performing standard survival analyses and calculating descriptive statistics (see below for further details).

Further rules were developed to confirm the optimum threshold and clustering time-interval, to maximise the sensitivity and specificity for outcome detection. Analysis included: identifying all potential prostate cancer-related events of interest (table 52); removing early confounding trial events (initial outcome-free periods), removing overlapping codes (where events were counted more than once), assessment of inclusion or exclusion of routine follow-up oncology appointments and assessment of

continual progressive states, which later defined our analysis to the first two progressive outcomes. The rules can be seen in table 51.

Model development rules (for first and subsequent events)	
1	Outcome triggered = when the number of prostate cancer-related events reached a defined threshold [^]
2	8-weekly clustering intervals [^]
3	Sustained increase above the threshold = the same outcome
4	Subsequent outcomes: When the number of events reduced then raised above ≥ 5 or ≥ 6 events again
5	An SRE could flag an outcome even if the threshold was not reached
6	SREs in the outcome peak = the same outcome until the threshold was reached again
Final algorithm training and validation rules (enhancing identification of 1st event)	
1	Outcome triggered = when the number of prostate cancer-related events reached the threshold (≥ 5) *
2	8-weekly clustering intervals
3	Sustained increase above the threshold = the same outcome
4	Subsequent outcomes: When the number of events reduced to ≤ 4 , then raised above ≥ 5 events *
5	An SRE could flag an outcome even if the threshold was not reached
6	SREs in the outcome peak = the same outcome until the threshold was reached again
7	Remove outcome if identified in the first 15% of a patient's history (likely to be trial treatments) *
8	Radical radiotherapy should not flag an outcome, as a definitive Tx for prostate cancer *
9	Routine outpatient appointment events included *

Table 51: The rules developed to detect the routine data-derived outcomes.

[^] = during model development the optimum thresholds required to trigger an event were assessed as ≥ 5 or ≥ 6 prostate-cancer related events per 8-week interval; * = amended or new rule.

Included events		
4	Chemotherapy	Chemotherapy assessment, blood sample, procurement and delivery, plus further potential chemotherapy codes
5	Imaging	Location specific, computerised tomography (CT), positron emission tomography (PET), magnetic resonance imaging (MRI), bone scan, X-ray, ultrasound, cystoscope, other imaging to specific areas, radiopharmaceutical imaging, non-prostate cancer scans (end-of-life care scans), scans related to SREs
6	Radiotherapy	RT planning, delivery, type, brachytherapy planning and delivery
7	Biopsy	Biopsy
8	Surgery	Surgery to prostate, bladder, lymph nodes, surgery for spinal cord compression and pathological fracture
9	Other therapies, interventions	Hormone therapy, immunotherapy, other drugs, blood transfusion including various interventions such as catheter insertion
10	SRE direct detection	SREs identified individually - Radiotherapy, surgery to bone, spinal cord compression and pathological fracture (identified through diagnoses and treatment codes)

Table 52: The prostate cancer-related interactions of interest, for the algorithm to detect in the routine data, to flag an outcome of interest.

Once the algorithm had been trained, validation could be undertaken, firstly, at a *patient surrogacy* level. During the single-site validation (cohort three, N=46), the number of STAMPEDE trial outcomes that were detected using the HES-algorithm were

calculated and the time to the first algorithm-detected outcome (AFS) per patient was compared to the STAMPEDE endpoints (FFS, PFS, MFS). Detection accuracy was calculated using the magnitude of the difference in weeks. In this cohort (N=46), the false positive and false negative events were confirmed by note review. The total number of events detected using each endpoint (HES AFS and SREs, STAMPEDE MFS, PFS, FFS and SREs) (N=93) were also analysed and compared.

To validate at a *trial surrogacy* level (cohort two onwards) the algorithm was reapplied to the routine data which was split by those receiving docetaxel (arms C, E) and those not (arms A, B, D, F; in the final validation stage, only arm A was utilised). Standard survival analyses (Kaplan-Meier) methods were used and Cox-proportional hazard models with 95% confidence intervals, to calculate the hazard ratios. This was to estimate and compare the treatment effects between the STAMPEDE outcomes (FFS, PFS, MFS and SREs), the clinical reference-derived outcome (progression) (where possible, during the single-site studies) and the routine data-derived outcome (AFS and SREs). These were defined as the time from trial randomisation to the first outcome per data source. The effect of grouping the outcomes into the 8-week intervals were also investigated, by assigning the STAMPEDE MFS into the routine data intervals, prior to re-calculating the STAMPEDE MFS hazard ratio (HR). A HR below 1.00 favoured the research group (chemotherapy - docetaxel), as opposed to the control standard-of-care (non-chemotherapy hormone therapy, all patients were randomised prior to docetaxel becoming added to the standard of care). The median survival (in weeks) and the 2-year (104 weeks) and 5-year (261 weeks) proportion remaining estimates (%) were also calculated.

In addition, during the multi-site validation (cohort four, N=1,695), Pearson's correlation coefficients were calculated comparing the correlation of the algorithm endpoints (AFS and SREs) to the STAMPEDE MFS, PFS, FFS and SREs. The strength of correlation was defined as detailed in the literature (199, 200) (for example, 0.00-0.30 = negligible correlation; 0.30-0.50 = low correlation; 0.5-0.70 = moderate correlation; 0.70-0.90 = high correlation; 0.90-1.00 = very high correlation) (199, 200). Correlation coefficients were calculated for all endpoints for outcomes where patients experienced both a HES and a trial outcome. In addition, using one example (the STAMPEDE MFS), further

correlation coefficients were calculated: those where patients only experienced a HES outcome (but not necessarily an MFS outcome) and those where *all* HES and STAMPEDE MFS outcomes were included (including those censored without an outcome). Both of these measures included patients censored at mutual time points due to not experiencing an event. The initial correlation coefficient calculated for outcomes identified in both sources (HES and the trial) was undertaken to gain an unbiased coefficient. For the other measures, censoring would lead to a mutual outcome date being identified, biasing the results. Hence, these results were deemed inconsequential.

Scatter graphs were created where appropriate to visualise the correlation, the line of perfect correlation was also present. Bland-Altman plots were also constructed to determine the quality of surrogacy for the HES outcome to the trial MFS and PFS. Limits of agreements were defined as the mean difference in the number of weeks between the HES and the STAMPEDE MFS ± 1.96 the standard deviation of the differences.

4.10 Results

4.10.1 Participants

Patient characteristics, for each cohort (1-4) (table 46) are shown in table 53.

		C1: Model development (N=6)	C2: Algorithm training (N=44)	C3: Algorithm validation (N=46)	C3: Algorithm validation (N=93)	C4: Algorithm validation (N=1,695)
		Number of patients (%)				
STAMPEDE treatment arm	A	2 (33%)	9 (20%)	16 (35%)	27 (29%)	855 (50%)
	B	2 (33%)	8 (18%)	6 (13%)	14 (15%)	-
	C	0 (0%)	6 (14%)	9 (20%)	15 (16%)	415 (24%)
	D	0 (0%)	6 (14%)	3 (7%)	9 (10%)	-
	E	1 (17%)	9 (20%)	6 (13%)	16 (17%)	425 (25%)
	F	1 (17%)	6 (14%)	6 (13%)	12 (13%)	-
Broad disease grouping	Newly diagnosed N0M0	0 (0%)	10 (23%)	7 (15%)	17 (18%)	343 (20%)
	Newly diagnosed N+M0	0 (0%)	3 (7%)	6 (13%)	9 (10%)	229 (14%)
	Newly diagnosed M1	6 (100%)	30 (68%)	32 (70%)	65 (70%)	1035 (61%)
	Previously treated M0	0 (0%)	1 (2%)	0 (0%)	1 (1%)	44 (3%)
	Previously treated M1	0 (0%)	0 (0%)	1 (2%)	1 (1%)	44 (3%)
Age at randomisation	Median (IQR)	60 (56-66)	63 (57-69)	65 (57-73)	63 (57-70)	66 (61 – 71)
	Range	54-68	48-80	41-81	41-81	42 - 82
PSA at randomisation	Median (IQR)	129 (92-207)	109 (57-348)	113 (23-409)	112 (41-406)	66 (21 – 189)
	Range	19-478	10-5000	3-8028	3-8028	0.2 - 15747
T category at randomisation	T0	0 (0%)	0 (0%)	0 (0%)	0 (0%)	6 (0%)
	T1	0 (0%)	0 (0%)	2 (4%)	2 (2%)	16 (1%)
	T2	0 (0%)	4 (9%)	4 (9%)	8 (9%)	175 (10%)
	T3	3 (50%)	26 (59%)	21 (46%)	49 (53%)	1095 (65%)
	T4	2 (33%)	7 (16%)	7 (15%)	14 (15%)	282 (17%)
	TX	1 (17%)	7 (16%)	12 (26%)	20 (22%)	121 (7%)
N category at randomisation	N0	2 (33%)	13 (30%)	14 (30%)	29 (31%)	730 (43%)
	N+	4 (67%)	21 (48%)	23 (50%)	45 (48%)	867 (51%)
	NX	0 (0%)	10 (23%)	9 (20%)	19 (20%)	98 (6%)
Metastases at randomisation	M0	1 (17%)	20 (45%)	19 (41%)	40 (43%)	616 (36%)
	M1	5 (83%)	24 (55%)	27 (59%)	53 (57%)	1079 (64%)

Table 53: Patient characteristics, split by the analyses cohort (cohort 1 model development N=6; cohort 2 algorithm training N=44; cohort 3 algorithm validation N=46 & 93; cohort 4 algorithm validation N=1,695).

C: cohort; TX: tumour cannot be measured; NX: nodal status cannot be measured; N+: node positive; N0: node negative. M1: metastatic; M0: non-metastatic. Due to rounding, percentages (%) may not add to 100%.

4.10.2 Model development (pilot/feasibility) (cohort 1)

58.6% (136/232) of all the prostate cancer-related interactions/events were identified in the routine data records, when compared to the clinical noting (range; 33.3%, 19/57 –

73.4%, 58/79). When inferring missing routine data records with alternative coding, the accuracy increased to 62.5% (145/232) and when assessing the accuracy for mandatory coded events only, the accuracy increased to 76.8% (129/168) (table 54). Therefore, 23.2% (39/168) of events were missing due to erroneous data.

	Patient				Total events
	1	2	3	4	
Missing event analyses					
Number of events identified in the HES	47	58	12	19	136
Total number of events (reference)	77	79	19	57	232
Percentage of events detected (%)	61.0	73.4	63.2	33.3	58.6
Inferred event analyses					
Number of events identified in the HES	51	59	12	23	145
Total number of events (reference)	77	79	19	57	232
Percentage of events detected (%)	66.2	74.7	63.2	40.4	62.5
Excluded event analyses					
Number of events identified in the HES	45	50	12	22	129
Total number of events (reference)	61	53	12	42	168
Percentage of events detected (%)	73.8	94.3	100.0	52.4	76.8

Table 54: The accuracy of the HES routine data coding for prostate cancer-related events, since randomisation, for the three different accuracy calculations (missing events, inferred events and excluded events) for four patients.

Missing event analyses = assessing all prostate cancer related records; inferred event analyses = assessing records that can be identified using coding infernal; excluded event analyses = assessing the impact of coding that was planned to be excluded (see table 49 for further details).

Graphically grouping/clustering the prostate cancer-related events, sourced from the note review data, optimally identified progression outcomes in 8-week intervals (identifying 11/11, sensitivity: 1.00, (table 55)).

Patient	Reference progression outcomes	Detected with model	% of outcomes detected
1	3	3	100
2	3	3	100
3	1	1	100
4	2	2	100
5	2	2	100
Total	11	11	100%

Table 55: The number of progression outcomes identified in the clinical noting that were detected using the clustering of events model. The patient numbers do not necessarily correspond to those in table 54.

For example, in figure 29, from analysing the case history, three progression outcomes were identified which reflected the completion of the two STAMPEDE CRFs (week 24, month 24), plus another CRF not completed (for progression 3); only the 1st instance has to be documented by the trial. The three progressions can be seen as three peaks (figure 29), with the highest point reaching week [16-24), [72-80) and [112-120). The patient developed increasing bone pain and a bone scan (week 16) confirmed progression, alongside a PSA increase (biochemical failure). This was noted in the STAMPEDE CRF as objective, biochemical, symptomatic and SRE progression. Subsequently another chemotherapy trial was initiated but the pain continued to worsen; the bone scan (week 66), confirmed objective progression. This led to cessation of the trial treatment. The pain again continued to worsen, and the PSA indicated progression, which was confirmed on a bone scan (week 102). A CRF (month 24) was completed for biochemical, objective and symptomatic progression, subsequently second line chemotherapy was initiated. The patient continued to progress further without any treatment response. In addition, multiple SREs were experienced throughout the disease history.

The events that were identified and included in the model as prostate cancer-related, can be seen in table 49. The analyses validating the 8-week intervals can be seen in the appendix (see 8.2.3, table 82, table 83).

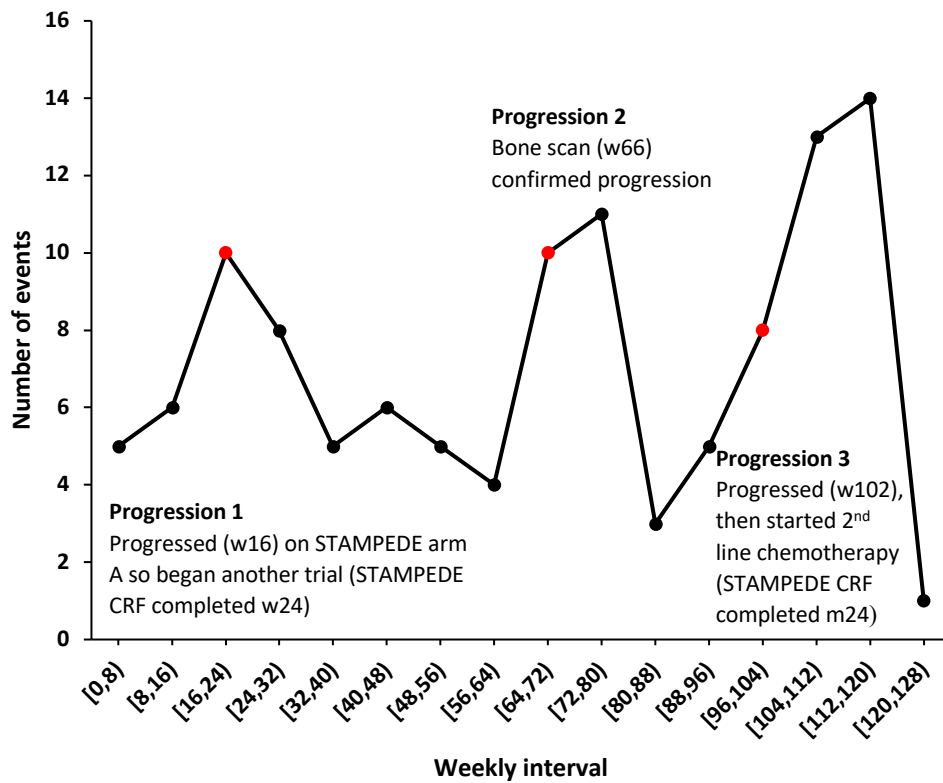


Figure 29: An example of the 8-weekly clustering of the note review sourced prostate cancer-related events, for one patient, from randomisation (t=0) until death.

Red circles = progression outcomes; '[' = inclusive of nearest value; ')' = exclusive of nearest value; w = week; # = number.

As shown, clustering the note review sourced prostate cancer-related interactions, enabled the identification of major trial outcomes (progression). Next the model had to be tested on the routine data, despite the data quality limitations highlighted in table 54. When this model was applied to the routine data, progression outcomes were also identified by the peaks of activity (figure 30).

Next a threshold had to be chosen, in which the number of prostate cancer-related events needed to reach, to flag an outcome of interest. Whilst considering the number of true and false positives, the optimum number of prostate cancer-related events in which to exceed (threshold) to identify an outcome per 8-week interval (in the five patients) were: ≥ 6 (identifying 12/13 outcomes, sensitivity: 0.92, PPV: 0.75) and ≥ 5 , (identifying 13/13 outcomes, sensitivity: 1.00, PPV: 0.72) (table 56) (figure 30). Further detail into the number of events detected, by patient, at both the five and six event thresholds can be seen in the appendix 8.2.4.

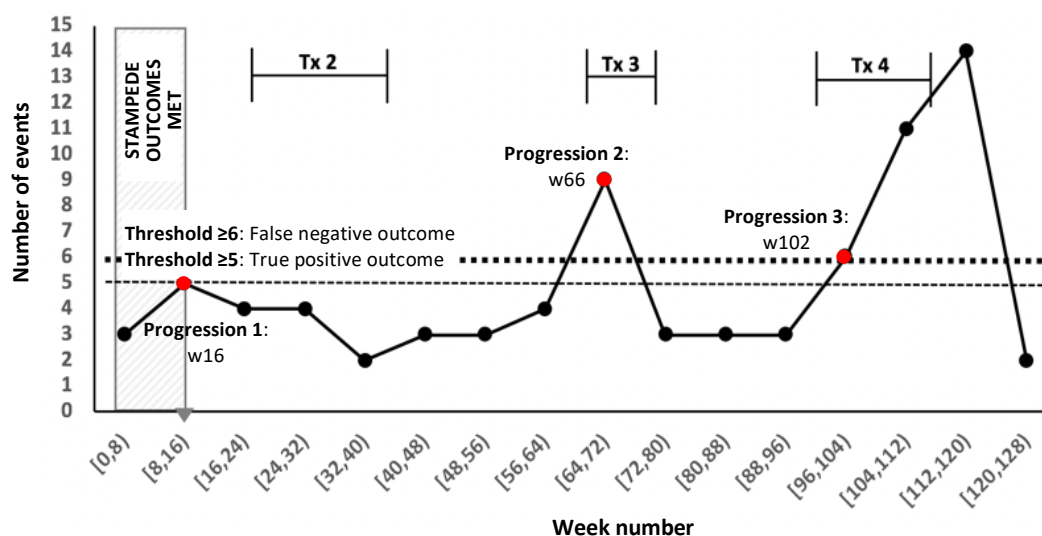


Figure 30: Clustering of the routine HES prostate cancer-related events (for the patient in figure 29) to identify outcomes of interest (progression = red dots).

The dashed lines correspond to the two optimum thresholds (≥ 5 and ≥ 6) required to trigger that an outcome of interest was met; '[' = inclusive of nearest value; ')' = exclusive of nearest value; w = week; Tx = treatment.

Threshold events	True positives	False positives	False positives (% extra events)	Sensitivity	False negatives	PPV
≥ 3	13	8	0.62*	1.00	0	61.9
≥ 4	13	6	0.46*	1.00	0	68.4
$\geq 5^{\wedge}$	13	5	0.38	1.00	0	72.2
$\geq 6^{\wedge}$	12	4	0.31	0.92	1	75.0
≥ 7	10	2	0.15	0.77*	3	83.3
≥ 8	9	2	0.15	0.69*	4	81.8
≥ 9	8	2	0.15	0.62*	5	80.0
≥ 10	7	1	0.08	0.54*	6	87.5
≥ 11	5	0	0.00	0.38*	8	100.0
≥ 12	2	0	0.00	0.15*	11	100.0

Table 56: Analyses into the optimum threshold to identify outcomes of interest. The threshold was chosen as a compromise between sensitivity and PPV. Specificity values were not possible to calculate, nor the NPV, due to an infinite number of false negatives.

(*) = the number of false positives were assessed to be too high, or the number of true positives too low to be clinically useable; (\wedge) = the two intervals chosen as a compromise between sensitivity and PPV. The thresholds are demonstrated in figure 30. The number detected, by patient, at the chosen thresholds (5, 6) can be seen in the appendix, table 84.

The three data sources (trial data, clinical noting and routine data) were triangulated to compare the ability to identify the outcomes (figure 31). At both thresholds (≥ 5 and ≥ 6 events), all outcomes identified in the clinical noting were identified in either the trial

data or by using routine data model (or both). False positives were identified; the trial data identified one outcome not present in the other data sources and the routine data identified four or five outcomes not present in the other sources (depending on the threshold).

The trial documented 9/13 of the outcomes detected using the clinical noting (sensitivity: 0.69, PPV: 0.90). At both thresholds, the 4/13 (30.8%) events not documented by the trial were identified using the routine data model (figure 31) (appendix, table 85).

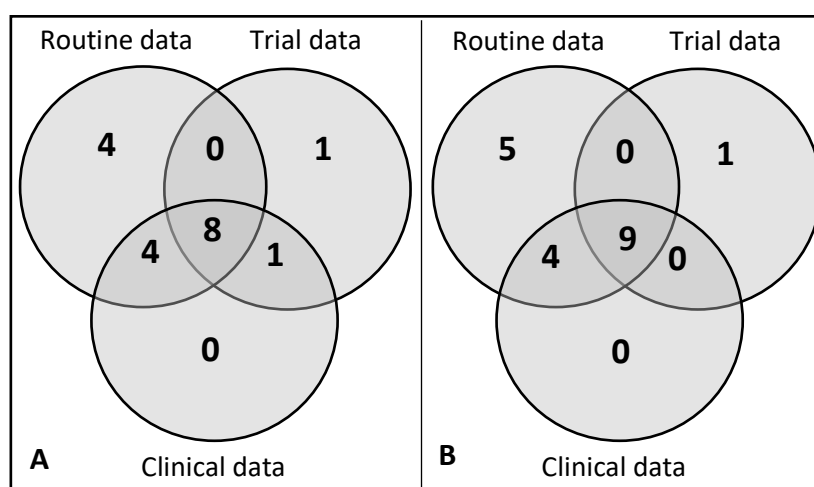


Figure 31: The number of progression outcomes identified by the data source at, (A) threshold ≥ 6 events and (B) threshold ≥ 5 events, corresponding to the data in table 56.

The mean difference in detection (mean magnitude of the difference) between the routine data identified outcomes and the clinical noting was 4.5 weeks, at both thresholds (range; -4 to +14 weeks) (appendix, table 86). At both thresholds, the model identified 8/13 outcomes within the same 8-week interval, 2/13 within one interval (± 8 weeks) and 2/13 within two intervals (± 16 weeks). The extra remaining outcome (1/13) identified by the ≥ 5 threshold (but not using the ≥ 6 threshold) was also detected within one interval (± 8 weeks) (appendix, table 85) (appendix, table 86). The analysis of the thirteen events can be seen in the appendix 8.2.5.

20 SREs were identified during note review, independent of thresholds (figure 32); 10/20 (sensitivity: 0.50) SREs were documented in the trial data. 16/20 (sensitivity: 0.80)

SREs were identified using the routine data model (after inferral, by identifying further radiotherapy sessions using clinical oncology indicators) (figure 32) (appendix, table 87).

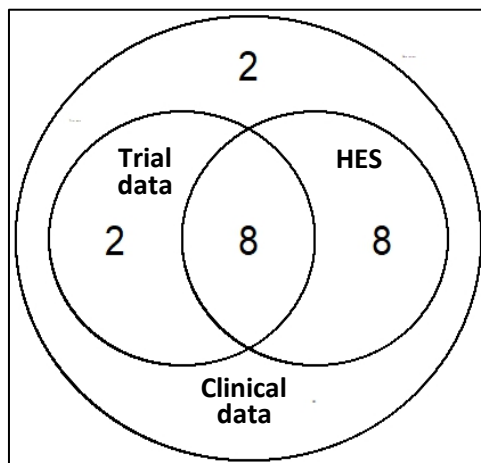


Figure 32: A Venn diagram showing the number of SREs identified in the trial data, the note review (clinical data) and the routine data.

The results are shown when inferring missing routine data radiotherapy events with routine data identified clinical oncology outpatient visits.

When splitting the patients into subgroups, by time (the earliest and latest patients randomised into the STAMPEDE trial), the sensitivity of detecting events increased from 0.69 (detecting 9/13 SREs) to 1.00 (detecting 7/7 SREs) (table 57) (see appendix: 8.2.6).

Data source	Total number of SREs	Number of SREs (3 earliest patients)	Number of SREs (3 latest patients)
Reference data			
Note review (reference)	20	13	7
Trial data	10	7	3
Routine data			
HES (without inferral)	8	4	4
HES (with inferral)	16	9	7
HES sensitivity (HES vs. note review)			
HES (without inferral)	0.40	0.31	0.57
HES (with inferral)	0.80	0.69	1.00

Table 57: Accuracy of the routine data and the trial data for identifying SREs, in comparison to the note review reference, split by date randomised. The number of SREs correspond to figure 32.

Inferral = enabling clinical oncology outpatient events to flag radiotherapy regimens.

At the end of the pilot study, the optimum rules to detect outcomes of interest were determined to be: clustering events into 8-week intervals, including SREs to flag

additional outcomes plus a threshold of five events. These were further tested during algorithm training.

4.10.3 Algorithm training (cohort 2)

Further sensitivity analyses (cohort 2, N=44) were undertaken to identify the optimum number of true positive progression outcomes, with the lowest number of false positives (prior to the first outcome (the AFS)). The 8-week clustering interval, with the threshold of ≥ 5 events was confirmed. In addition, a threshold of ≥ 4 events in an interval was required to be reached before a subsequent outcome could be triggered (or the peak would be classified as the same outcome). The inclusion of oncology outpatient events was also confirmed to enhance outcome detection and removal of the first 15% of events per patient was required to remove confounding trial randomisation activity.

At an individual trial *outcome* level (assessing *patient level surrogacy*), these optimum rules identified the majority of the first two progression outcomes or classified that no outcome had occurred for 59/64 (92.2%) outcomes. In the 53 patients that progressed, 42/53 (sensitivity: 0.79) of the first two progression outcomes were identified, with three false positives prior to the first. When restricting analyses to the first progression only (as the time to the first outcome was proposed to be used as the trial endpoint, the AFS), 34/44 (77%) were identified in any time interval (progressed or did not progress). However, upon restricting detection to within 16 weeks (≤ 2 intervals), fewer progressions were identified, 28/44 (63.6%). At an individual *patient* level, the algorithm correctly classified the majority of patient's disease states as, 1) progressed (31/33, sensitivity: 0.93) or 2) did not progress (9/11, specificity: 0.82). The reasons for the delayed (6/44) and missed (10/44) progressions can be seen in table 58.

Description	Number
False negatives	
Non-trial events were not coded (missed), for example, radiotherapy	1
Hormone therapy was initiated/continued and was not in the HES	6
The number of events did not reach the threshold to trigger an outcome	1
Trial therapy was initiated and was not in the HES	2
Delay >2 intervals	
There was a delay in starting treatment for progression	2
Hormone therapy was initiated/continued prior to other treatment/scans	3
The progression was identified clinically at a delay	1
Trial therapy was initiated prior to other scans/treatments	2
False positives	
Detected death from a non-prostate cancer cause (although constituting the trial MFS)	1
Investigation for a suspected progression but progression was not confirmed	1

Table 58: Reasons for the algorithm-missed/delayed outcomes.

Sometimes an outcome was missed/delayed for more than one reason.

When comparing the first routine data outcome to the documented STAMPEDE endpoints (MFS, PFS and FFS), a mixture of the traditional endpoints were being identified. Within any time-interval, the routine data algorithm identified 10/32 FFS (31.3%), 13/24 PFS (54.2%) and 15/25 MFS (60%) of the STAMPEDE outcomes, with a mean magnitude of the difference of two intervals (9.7 weeks) (range; -8 to +44 weeks). Restricting the acceptable delay in detection reduced the number of endpoints that were detected.

In addition, four outcomes not documented by the trial, were identified by the HES model (confirmed via note review). Three of these outcomes were missed from the trial data erroneously. Of these, the HES model identified one outcome one week early and one outcome two weeks late (compared to the date of objective progression in the clinical noting). The third was identified 90 weeks late, once chemotherapy was initiated. The fourth outcome not documented by the trial was excluded from the subsequent analysis due to the censoring (the progression outcome occurred in the same week as the STAMPEDE censor) and hence was not yet documented in the trial data. Excluding these four events, 29 patients fulfilled at least one STAMPEDE endpoint and the routine data algorithm identified 21/29 (72.4%).

The total number of outcomes detected using the HES algorithm (27 events) was most comparable to the number of STAMPEDE MFS (25 events) and PFS (24 events) outcomes identified (table 59). The number of SREs detected *directly* using the HES data (7 events)

were not comparable to the number detected in the STAMPEDE data (18 events) (table 59).

Data/Outcome	Total outcomes detected
Note review PFS	33
STAMPEDE PFS	24
STAMPEDE MFS	25
STAMPEDE FFS	32
HES AFS	27
STAMPEDE SRE	18
HES SRE	7

Table 59: The number of outcomes identified during the note review (note review PFS), those documented in the trial (STAMPEDE PFS, MFS, FFS, SRE) and those identified using the HES algorithm (HES AFS, SRE), from randomisation, until the mutual censor date.

See the appendix table 88 for the number of outcomes detected, split by treatment.

Endpoint treatment effects were then investigated for *trial level surrogacy*. The routine data algorithm was capable of separating the docetaxel treatment benefits in line with the standard trial endpoints and the clinical reference data (table 60) (figure 33). The routine data hazard ratio was most similar to the trial PFS, with both hazard ratios at 0.41 (HES AFS 0.41, 95% CI 0.16-1.02; STAMPEDE PFS 0.41, 95% CI 0.15-1.10) (table 60). An example from the time to event analyses can be seen in figure 33 (HES vs. the STAMPEDE MFS). The number of events that were identified for each endpoint, by the trial treatment, can be seen in the appendix (8.2.7).

Data/Outcome	Hazard ratio (95% CI)
Note review PFS	0.51 (0.23 - 1.13)
STAMPEDE PFS	0.41 (0.15 - 1.10)
STAMPEDE MFS	0.46 (0.18 - 1.16)
STAMPEDE FFS	0.47 (0.21 - 1.01)
HES AFS	0.41 (0.16 - 1.02)
SREs	
STAMPEDE SRE	0.72 (0.25 - 2.05)
HES SRE	0.58 (0.11 - 3.00)

Table 60: The hazard ratios and 95% CIs comparing the HES model output for the HES AFS and SREs, compared to the trial-derived endpoints (FFS, PFS, MFS and SRE) and the note review derived progression (note review PFS).

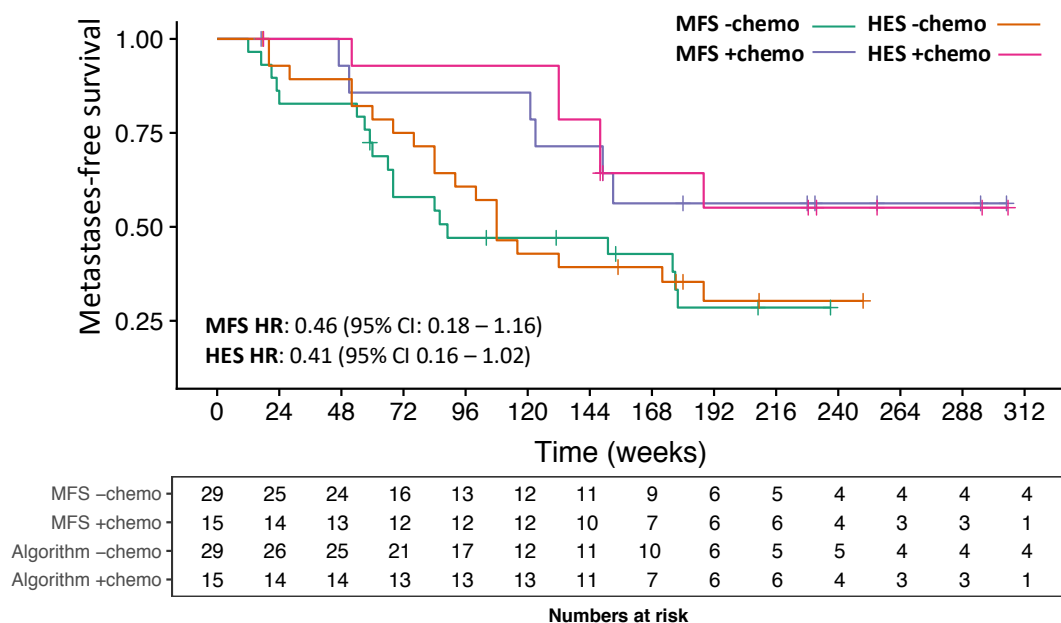


Figure 33: Kaplan-Meier analyses for STAMPEDE chemotherapy (C, E) and non-chemotherapy (A, B, D, F) treatment arms, comparing the algorithm output with the STAMPEDE trial MFS. Numbers at risk are shown (including those remaining = not censored or no outcome of interest reached).

The median survival time using the routine data algorithm (HES AFS) for the non-chemotherapy arms was 108 weeks; this value was most comparable to the STAMPEDE MFS and PFS (92 and 89 weeks). The routine data algorithm (HES AFS) median was not reached for the chemotherapy arms. The STAMPEDE PFS did not reach the median either (table 61).

The 2-year event free rate using the routine data algorithm (HES AFS), for non-chemotherapy arms was 57%, most similar to the STAMPEDE PFS (49%). The 2-year event free rate for the chemotherapy arms using the HES AFS was 93%, most similar to the MFS (79%) and PFS (79%). In addition, the 5-year event-free rates for the HES AFS were most comparable to the STAMPEDE PFS and MFS (Table 61).

Data	Endpoint	2y (104w) event-free rate (% remaining)	5y (261w) event-free rate (% remaining)	Median time to MFS (weeks)
Non-chemotherapy arms (ABDF)				
HES	AFS	57	30	108
STAMPEDE	PFS	49	30	89
	FFS	29	17	65
	MFS	45	32	92
Chemotherapy arms (CE)				
HES	AFS	93	55	NR
STAMPEDE	PFS	79	63	NR
	FFS	67	47	127
	MFS	79	59	321

Table 61: The median (weeks) and 2-year and 5-year event free rate (%) comparing chemotherapy and non-chemotherapy arms using the STAMPEDE endpoints and the HES-derived endpoint.

NR = not reached.

4.10.4 Algorithm validation

4.10.4.1 Small scale validation (N=46, N=93)

At a *patient surrogacy level*, the routine data algorithm identified 14/33 (42.4%) FFS, 23/29 (79.3%) PFS and 22/28 (78.6%) MFS endpoints. The routine data identified three events missed in STAMPEDE and identified one progression outcome that may not have been correct to document in the trial (a false positive routine data and trial-detected outcome). When removing these four patients from the analyses, 30/36 (83.3%) of one or more of the MFS, PFS or FFS endpoints were identified using the routine data model. Excluding false positive and negative STAMPEDE and HES-detected outcomes, there were 26 patients that experienced an outcome; of these, the mean delay in detection was 17.2 weeks.

Upon analysis of the cumulative cohort (N=93), despite the least FFS outcomes being identified using the routine data model, the total number of first outcomes (AFS) detected by the HES algorithm (66 events) were most comparable to the number of STAMPEDE FFS outcomes (67 events) (table 62). The number of SREs detected using the HES data (15 events) were again not comparable to the number detected in the STAMPEDE data (34 events) (table 62).

Data/Outcome	Total outcomes detected
STAMPEDE PFS	56
STAMPEDE MFS	56
STAMPEDE FFS	67
HES AFS	66
STAMPEDE SRE	34
HES SRE	15

Table 62: The number of outcomes documented in the trial (STAMPEDE PFS, MFS, FFS, SRE) and identified using the HES algorithm (HES AFS, SRE), from randomisation, until censor date.

See the appendix table 89 for the number of outcomes detected, split by treatment.

At a *treatment surrogacy level*, the routine data algorithm was again capable of separating the docetaxel treatment benefits in line with the traditional trial endpoints (figure 34) (table 63). The FFS hazard ratio, on this occasion, was the closest of the STAMPEDE endpoints to the HES-derived AFS endpoint (table 63). The number of events that were identified for each endpoint, by the trial treatment, can be seen in the appendix (8.2.7).

Data/Outcome	HR (95% CI)
STAMPEDE PFS	0.36 (0.18 - 0.72)
STAMPEDE MFS	0.47 (0.25 - 0.89)
STAMPEDE MFS (+ interval grouping)	0.47 (0.25 - 0.89)
STAMPEDE FFS	0.51 (0.29 - 0.89)
HES AFS	0.59 (0.33 - 1.03)
SREs	
STAMPEDE SRE	0.46 (0.20 - 1.05)
HES SRE	0.64 (0.20 - 2.00)

Table 63: The hazard ratios and 95% CIs comparing the HES model output for the HES-derived outcome and HES detected SREs, with the trial-derived endpoints (FFS, PFS, MFS and SREs).

To assess the impact of the time interval grouping on the routine data-derived outcome hazard ratios, the STAMPEDE trial data MFS outcomes were allocated to the corresponding intervals, prior to recalculating the hazard ratios. This did not change the STAMPEDE MFS hazard ratio, suggesting that the grouping of routine data events did not have significant implications on the results (figure 34, B) (table 63).

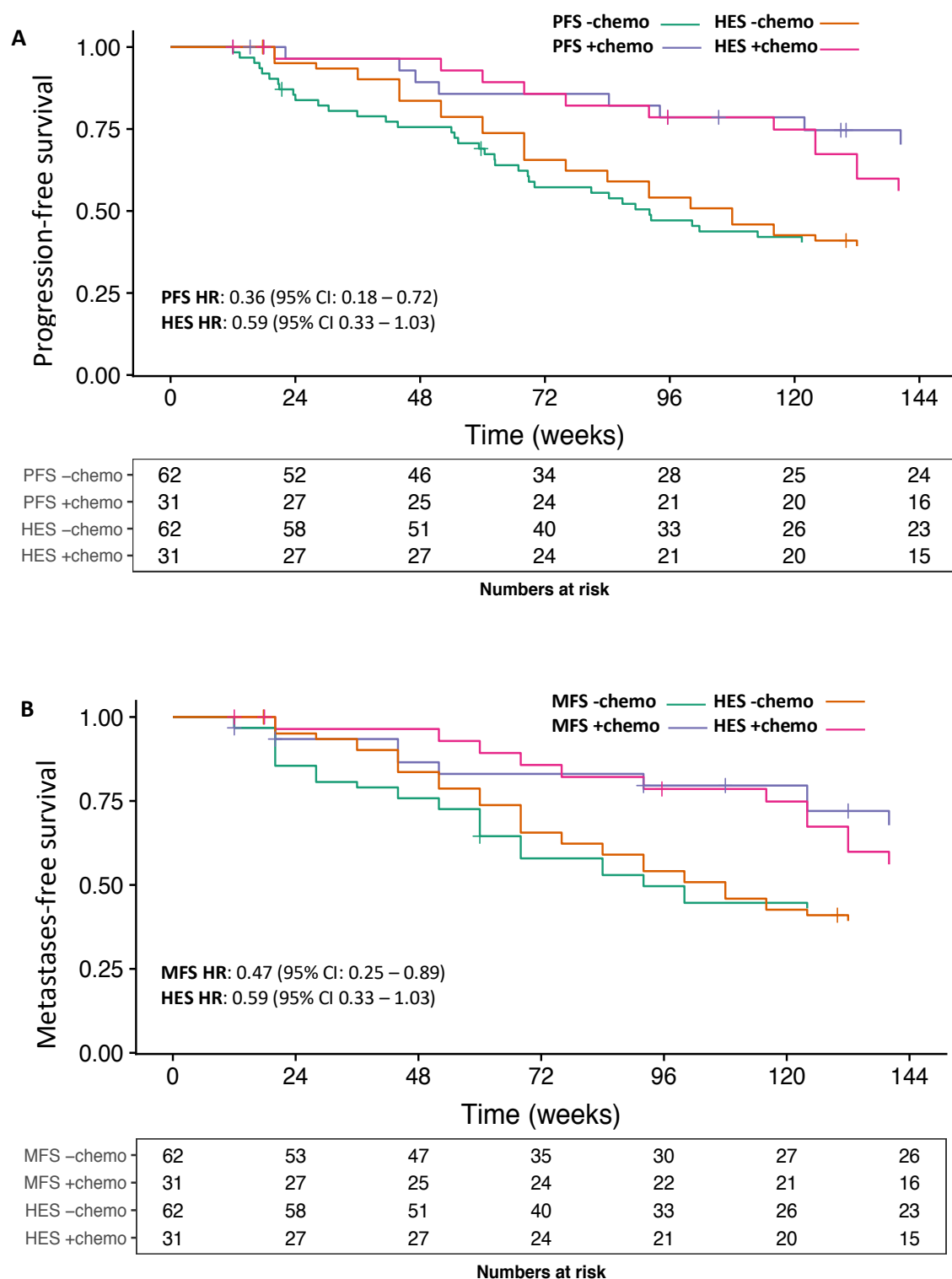


Figure 34: Kaplan-Meier analyses for STAMPEDE chemotherapy (C,E) and non-chemotherapy (A,B,D,F) treatments, comparing the routine data outputs with the (A) exact trial PFS and (B) with the trial MFS grouped into the corresponding time-intervals used in the algorithm.

Numbers at risk are shown (including those remaining = not censored or no outcome of interest reached).

The median survival times using the routine data algorithm for non-chemotherapy and chemotherapy arms were 108 and 148 weeks respectively. For the non-chemotherapy arms, this was most comparable to the STAMPEDE MFS and PFS (MFS, PFS: 92 weeks; HES AFS: 108 weeks). For the chemotherapy arms, the HES-derived estimate was most similar to the STAMPEDE FFS (FFS: 169 weeks; HES AFS: 148 weeks) (table 64).

The 2-year event free rate using the routine data algorithm (HES AFS), for non-chemotherapy arms was 51%, most similar to the STAMPEDE MFS (45%). For the chemotherapy arms, the HES-derived estimate rate was 79%, equal to the STAMPEDE PFS (79%). In addition, the 5-year event-free rates can also be seen (table 64).

Data	Endpoint	2y (104w) event-free rate (% remaining)	5y (261w) event-free rate (% remaining)	Median time to MFS (weeks)
Non-chemotherapy arms (ABDF)				
HES	AFS	51	26	108
STAMPEDE	PFS	44	28	92
	FFS	32	21	62
	MFS	45	32	92
Chemotherapy arms (CE)				
HES	AFS	79	39	148
STAMPEDE	PFS	79	65	321
	FFS	68	39	169
	MFS	59	59	321

Table 64: The median (weeks) and 2-year and 5-year survival (%) comparing chemotherapy and non-chemotherapy arms using the STAMPEDE endpoints and the HES-derived endpoint.

4.10.4.2 Large-scale validation (N=1,695)

The total number of progression outcomes detected using the HES algorithm (833 events) was most comparable to the number of STAMPEDE MFS (748 events) and PFS (754 events) outcomes (table 65). STAMPEDE identified more SREs (411 events), compared to the HES data (323 events) (table 65).

Data/Outcome	Total outcomes detected
STAMPEDE PFS	754
STAMPEDE MFS	748
STAMPEDE FFS	1028
HES AFS	833
STAMPEDE SRE	411
HES SRE	323

Table 65: The number of endpoint events documented in the trial (PFS, MFS, FFS, SRE) and identified using the HES algorithm (AFS, SRE), from randomisation, until censor date.

See the appendix table 90 for the number of outcomes detected, split by treatment.

The HES-derived endpoint was capable of showing treatment benefits across the multi-site cohort (N=1,645), with a hazard ratio closest to the STAMPEDE MFS (HES AFS HR: 0.88, 95% CI 0.77 – 1.01; STAMPEDE MFS HR: 0.82, 95% CI 0.71-0.95) (figure 35, A). The HES detected SRE hazard ratio was also comparable to the STAMPEDE SRE hazard ratio (HES SRE HR: 0.68, 95% CI 0.55-0.85; STAMPEDE SRE HR: 0.61, 95% CI 0.50-0.74) (table 66) (figure 35, B).

Data/Outcome	HR (95% CI)
STAMPEDE PFS	0.74 (0.64 - 0.86)
STAMPEDE MFS	0.82 (0.71 - 0.95)
STAMPEDE FFS	0.67 (0.60 - 0.76)
HES AFS	0.88 (0.77 – 1.01)
STAMPEDE SRE	0.61 (0.50 – 0.74)
HES SRE	0.68 (0.55 – 0.85)

Table 66: The hazard ratios and 95% CIs comparing the HES model output for the routine data derived-outcomes and HES identified SREs, compared to the trial-derived endpoints (STAMPEDE FFS, PFS, MFS and SREs).

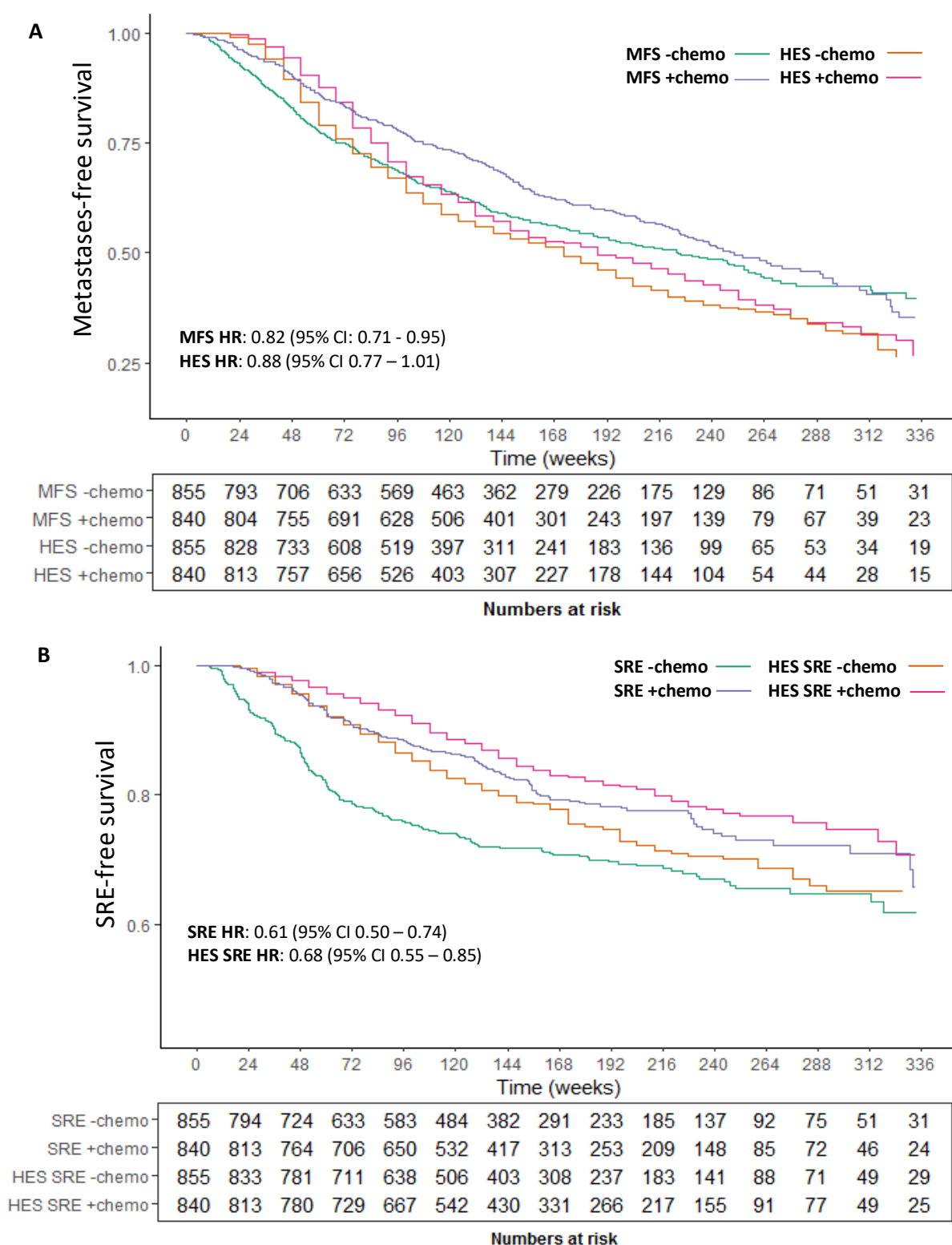


Figure 35: Kaplan-Meier analyses (without censor lines) comparing docetaxel (+chemo) with the standard-of-care arm (-chemo). The HES AFS and SRE endpoint are compared to, A) the STAMPEDE MFS, B) the STAMPEDE SREs. The steps in the HES curves = the clustering intervals.

The median (weeks) and 2-year and 5-year event free rate (% remaining) comparisons can be seen in table 67, as an example comparing the HES-derived and STAMPEDE MFS endpoints.

Endpoint	2y (104w) event-free rate (% remaining)	5y (261w) event-free rate (% remaining)	Median time to MFS (weeks)
Non-chemotherapy arms (A)			
HES AFS	64	37	172
STAMPEDE MFS	67	45	224
Chemotherapy arms (CE)			
HES AFS	67	38	188
STAMPEDE MFS	76	49	250

Table 67: The median (weeks) and 2-year and 5-year survival (%) comparing chemotherapy and non-chemotherapy arms using the STAMPEDE endpoints and the HES-derived endpoint.

To further investigate the relationships between the endpoints, the strength of the correlation between the HES-derived endpoints (AFS and SREs) and the STAMPEDE-derived endpoints were next compared (MFS, PFS, FFS and SREs). The correlation coefficients where both the HES and STAMPEDE endpoints identified an event can be seen in table 68. The coefficients between the HES-derived outcomes (AFS and SREs) and the STAMPEDE MFS, PFS and SREs showed *high* correlation (AFS vs. MFS = 0.74 (95% CI: 0.70 – 0.77); AFS vs. PFS = 0.73 (95% CI: 0.69 – 0.77); HES SREs vs. STAMPEDE SREs = 0.81 (95% CI: 0.75 – 0.86)) (table 68). In contrast, the HES AFS showed *low* correlation to the STAMPEDE FFS (table 68).

Data comparisons	Number of outcomes identified by both HES & STAMPEDE	Correlation coefficient (95% CI)
HES AFS vs. STAMPEDE MFS	565	0.74 (0.70 - 0.77)
HES AFS vs. STAMPEDE PFS	580	0.73 (0.69 - 0.77)
HES AFS vs. STAMPEDE FFS	699	0.67 (0.63 - 0.71)
SREs		
HES SREs vs. STAMPEDE SREs	178	0.81 (0.75 - 0.86)

Table 68: The correlation coefficients calculated to compare the HES endpoints to the standard STAMPEDE trial endpoints.

Correlation scatter plots comparing the HES-derived AFS outcomes and the STAMPEDE MFS and FFS (the highest and lowest correlation, respectively) can be seen in figure 36. The scatter plot comparing the HES and STAMPEDE SRE endpoints can be seen in the appendix, figure 49.

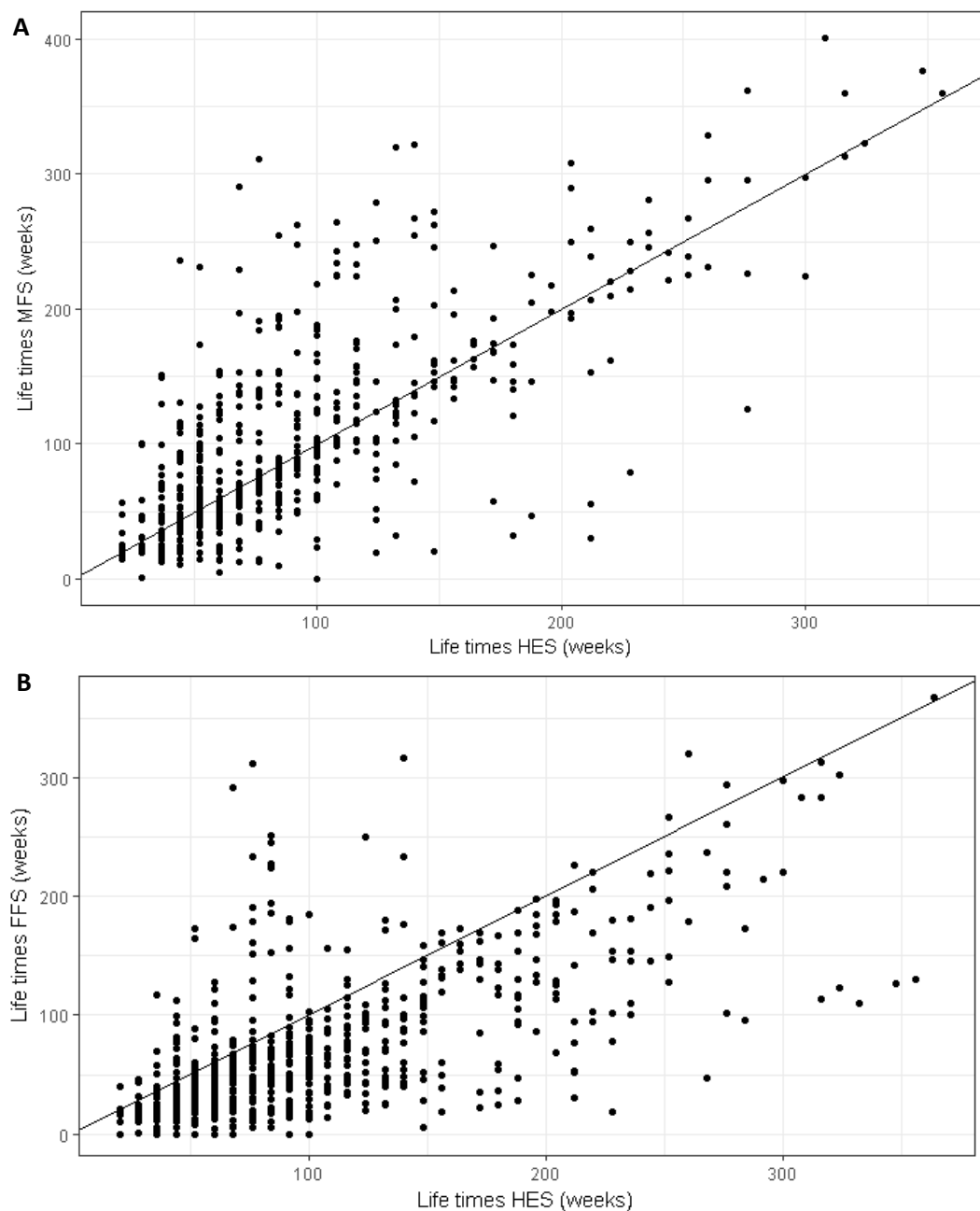


Figure 36: A scatter plot showing the correlation between the HES endpoint and the STAMPEDE, A) MFS and B) FFS.

Black line through the origin = a line illustrating 100% correlation. The vertical plotting of the lines is due to the 8-week clustering intervals on the x-axis.

A Bland-Altman plot, comparing the HES AFS and the STAMPEDE MFS and PFS outcomes can be seen in figure 37. The mean difference between the HES AFS and STAMPEDE MFS was -13 weeks. Hence, the HES outcomes were identified a mean of 13 weeks prior to the trial MFS. For the trial PFS, the mean difference was -3 weeks; hence, the HES outcomes were identified a mean of 3 weeks prior to the trial PFS.

The limits of agreement were wide at -112 and 87 weeks (HES AFS vs. MFS); it was therefore identified that 95% of the HES-detected outcomes were expected to lie 112 weeks before and 87 weeks after the STAMPEDE MFS (figure 37, A). These wide limits of agreement (199 weeks), may highlight the identification of a novel endpoint. When comparing the HES endpoint to the PFS, the limits of agreement were marginally narrower (187 weeks), at -97 and 90 weeks; 95% of the HES-detected outcomes were expected to lie 97 weeks before and 90 weeks after the MFS (figure 37, B). Again, potentially highlighting, that as with the MFS, a novel endpoint was being identified.

Plot patterns for both the MFS and PFS comparisons suggest that outcomes occurring in the first 100 weeks, generally lay within two standard deviations of the mean (figure 37). However, as time increased, a larger number of events lay outside the limits of agreement, suggesting that the HES endpoint may be more comparable to the trial endpoints for earlier events (occurring within the first two years of randomisation into the trial) (figure 37).

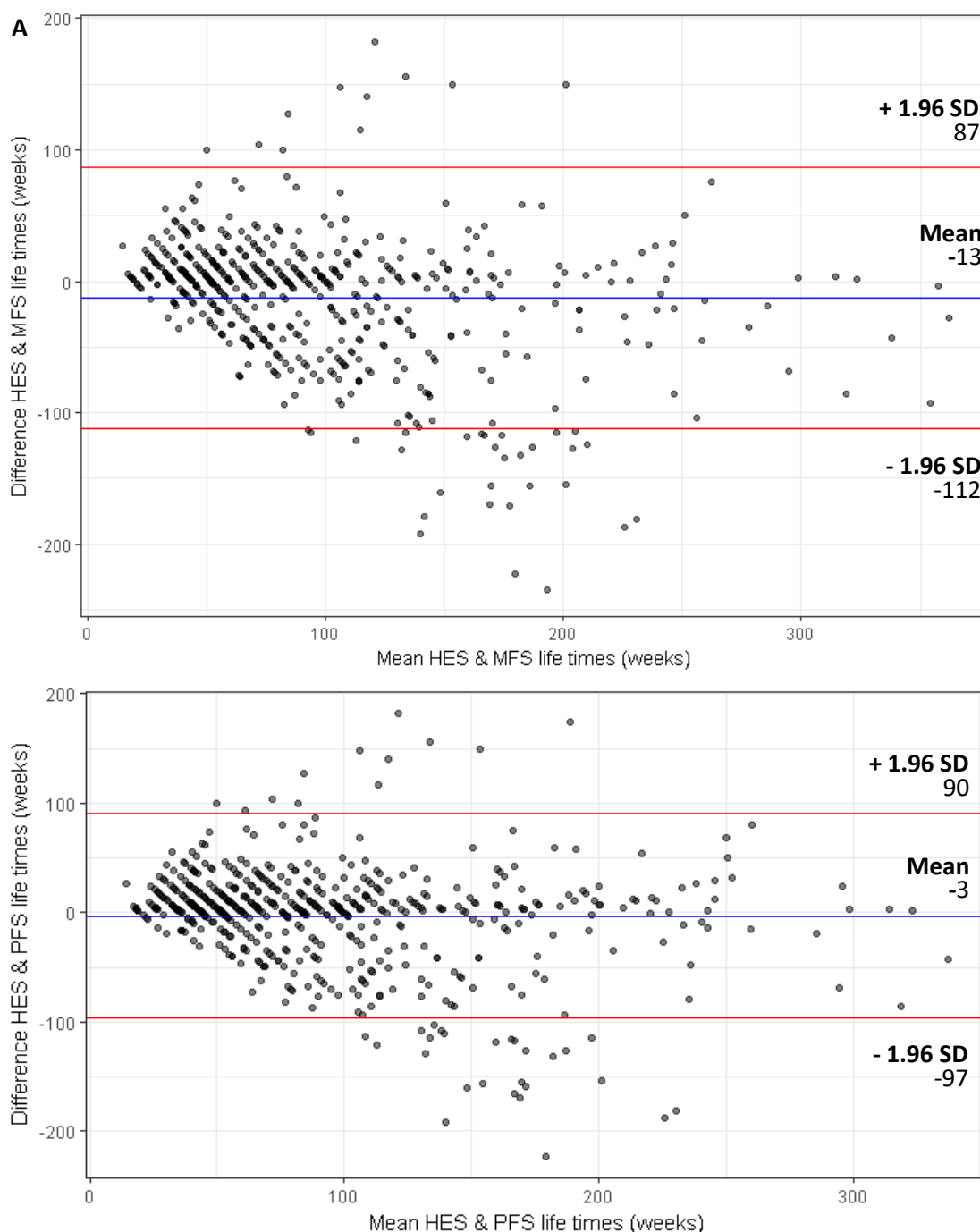


Figure 37: Bland-Altman plots between the HES AFS and the STAMPEDE A) MFS and B) PFS.

Blue line = the mean difference in detection of the HES and the STAMPEDE outcomes (weeks). Red lines = the limits of agreement, which are defined as the mean difference \pm 1.96 times the standard deviation (SD) of the difference between the HES and the STAMPEDE outcomes. Grey points = single outcomes; black points = where greater than one outcome overlaps.

For SRE detection, the mean difference between the HES detected endpoint and the STAMPEDE endpoint was 18 weeks. Hence, the HES events were identified a mean of 18 weeks after the trial SREs. The limits of agreement were at -53 and 89 weeks.

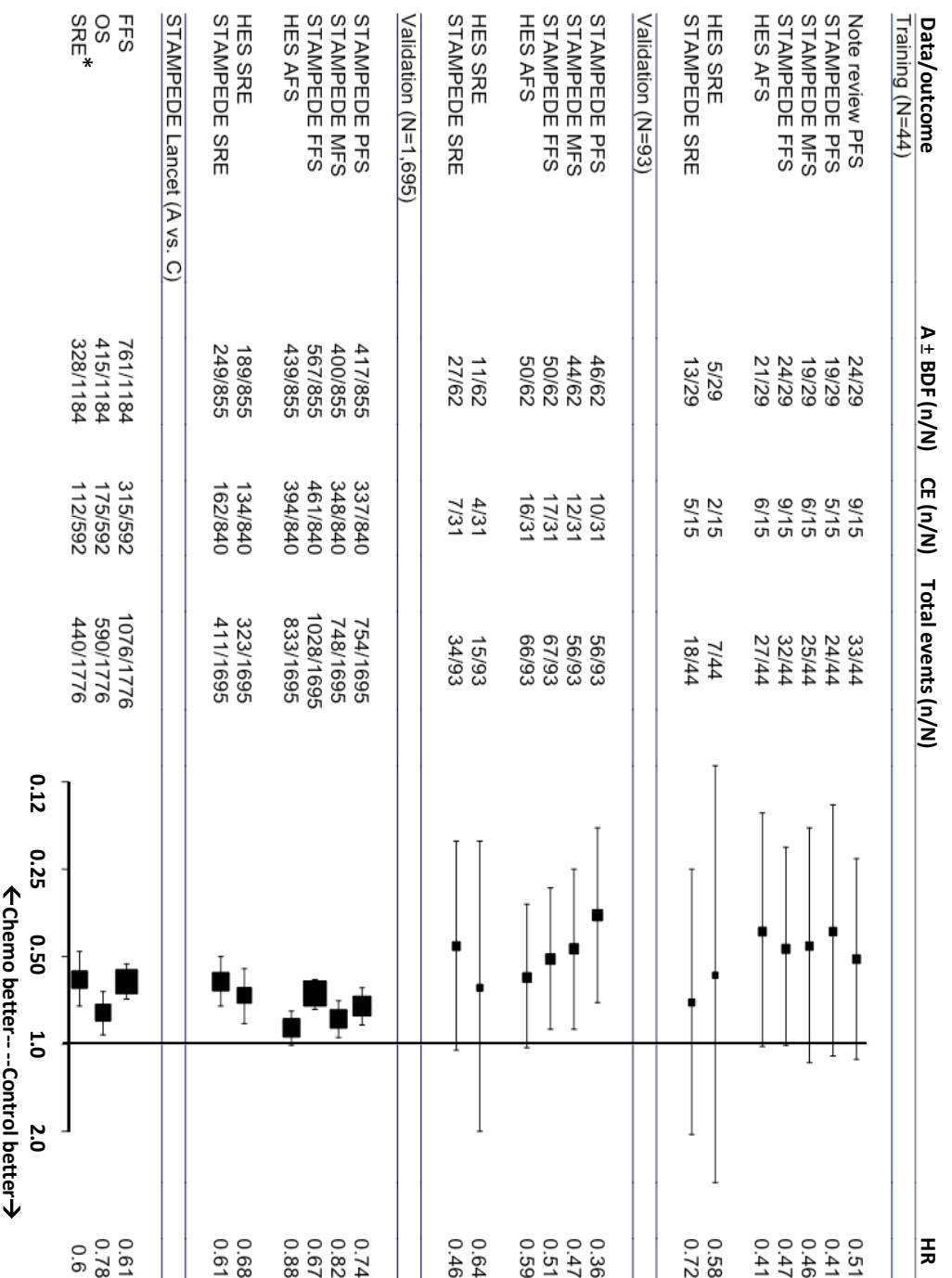


Figure 38: A summary forest plot. Hazard ratios (HR) (calculated using Cox-proportional hazards models), number of events (n), total number of patients (N) and confidence intervals for each comparison for chemotherapy (CE) and non-chemotherapy (A ± BDF) arms for the different cohorts are displayed.

* = 347/415 (84%) and 143/175 (82%) were prostate cancer related deaths.

4.11 Discussion

There is an increasing consensus that the use of routine healthcare records can aid the growing challenges of trial conduct and improve the quality of trials. The hypotheses is that this alternative trial follow-up design may: increase recruitment (63); reduce costs enabling funds and resources to be redistributed elsewhere (63, 103); reduce patient and trial staff burden (103); and subsequently enable patients more expedient access to novel treatments (201). However, as discussed and shown, there are well known concerns of missing data or inadequate outcome fields to enable this (201). Despite this, it has been proposed that routine data could be used to validate objective cancer progression and investigate subsequent treatments, explore progression characteristics and explore alternate progression endpoints (202). Routine data offers opportunities to develop endpoints for oncology trials, but practically, I found little evidence of previous attempts to identify non-survival endpoints and no evidence using HES data linked to a RCT.

Both administrative and registry data have the potential to identify outcomes (as discussed in the introduction, section 1.2.3). Registry data, despite often enabling collection of required outcomes of interest, often requires manual input for data collection. For example, during the completion of this PhD, the cancer outcomes and services dataset (COSD) has been developing its documentation of trial-related endpoints (such as, recurrence and progression); data were first collected in July 2015 and a new specification for enhanced collection was released in April 2018. However, these data on outcomes are manually generated at multiple disciplinary meetings (MDT) and thus relies upon clinician input, leading to missing data and loss to follow-up. Hence, I proposed that routine mandatory administrative data could, although not providing a *direct* source of outcomes, *indirectly* identify these. Hence, routine administrative data were the data source of choice during this chapter.

Standard trial non-survival endpoints (for example, FFS, MFS, PFS and SREs) present clinically different timepoints but still provide similar differences in intervention and function (to compare treatment efficacy) as surrogates for overall survival. However, new oncology trial surrogates are sought after (2) as existing surrogate outcomes are

proposed to be inadequate (203). This chapter set out to identify clinically useable non-survival endpoints using routine administrative healthcare data.

The three data sources (routine data, trial data and clinical noting data) each had their own limitations. For example, the clinical reference obtained during note review meant that events occurring outside of the hospital were often not accessible, early events were missing due to the transition from paper to electronic noting and wrongly recorded events may have been present due to incomplete or lost data (204). Hence, routine data events assumed to be false positives may indeed have been true positives. In the absence of the possibility to query back to the individual sites within this study, this was not possible to investigate. However, this study revealed initial evidence that events missed by the trial could be identified using routine data (for example, figure 31) and hence, it may suggest that routine data could be identifying further events that were missed in the clinical noting. Further to this, clinical note review was also extremely resource intensive and time consuming. This meant that a second reviewer could not audit my data collection process. This could mean that events thought to be true positives or negatives in the reference were not, impacting the algorithm accuracy statistics.

The trial data also had limitations; the STAMPEDE trial is only required to collect the first instance of events, for example, progression on treatment, so analyses of subsequent events is not possible, this is a common feature of trials. In addition, as illustrated in chapter 3, new hypotheses arise that were not previously appreciated, for example, response to subsequent treatments, and it is thus not traditionally possible to retrospectively collect these outcomes to enable the analyses. Trial data are also commonly known to have problems with loss to follow-up, where outcomes are missed. Hence, desirable characteristics of a routine data model would be to reduce loss to follow-up and enable analyses of events not routinely collected in the trial. The two data sources used as a reference, the clinical noting and the trial data were hence, not a true gold standard, as there is no known truth (147), making validation challenging.

As previously stated, assessment of the routine data within the other chapters (chapter 3 and 5) identified limitations, including, the erroneous coding and inadequate

availability of key outcome data fields (for example, progression). In addition, this study identified that some events are not required to be collected (not mandatory) in the administrative data due to the payment function of these data and thus cannot be analysed, for example X-rays. Again, this resulted in missing operational procedures and diagnostic codes (table 54).

Erroneously coded data were identified; for example, when SREs were not being detected (figure 32), which was also highlighted in the varying hazard ratios compared to the STAMPEDE SRE measure. Despite these missing data, in the largest analyses, when the effect of docetaxel on skeletal related events (SRE) was compared to the standard-of-care, the benefit of the intervention was still maintained using the HES-detected SRE measure (figure 35, B) (table 66). Although SRE events were shown to be missed from the routine data, events were identified that were not documented in the trial data, hence leading to the frameshifted but comparable hazard ratios seen (table 66). HES SREs were identified a mean of 18 weeks after the trial documented date. SREs are well defined events; for example, if a patient has had radiotherapy to the bone. Hence, this delay is implying that the HES data were often missing the trial detected events but may be identifying downstream SREs not collected in the trial. This hypothesis is supported as HES and STAMPEDE identified a different number of SREs (323 and 411 respectively) (table 65); with only 178 instances where an event was identified in both datasets. Hence, it is possible that the estimated correlation is high, as we are excluding events that were only reported by one dataset or are not correctly matched. This is difficult to assess as there is no known 'truth'.

For all the prostate cancer-related interactions, 23% of events were missed in the routine data when compared to the reference datasets (clinical noting and trial data). However, these were patients randomised earlier in the STAMPEDE trial and the SRE accuracy was seen to increase over time (table 57), in the very small sample. This was to be further investigated in a larger sample in chapter 5.

There is a continual drive for cleaner data to enhance reimbursement for the NHS (205) but cleaner data are also proposed to aid secondary analysis (206). In addition, it was seen that some procedures were not missed, however, were coded using alternative

descriptions. For example, radiotherapy sessions were often coded as clinical oncologist outpatient visits and scans were often coded as allied health professional visits, but the procedure was not documented. Hence, to detect as many events as possible, it was essential that these codes were also included in a model.

As predicted, the inadequate availability of key outcome data, missing and excluded coding, meant that these data could not be *directly* used for endpoint analyses. This meant that a model had to be created, based upon a set of rules, to *indirectly* infer when trial outcomes of interest occurred.

To account for these limitations, I developed a model using event clustering to flag absent outcomes from diagnosis and procedure patterns. This model was proposed to allow events to contribute, *indirectly*, even if they were coded using alternative descriptions. The model was designed to identify ‘clinically relevant outcomes’ that had been experienced. The aim was to identify the time to the first major clinical outcome. The hypothesis was, does an increase in HES activity correlate with clinically relevant outcomes? This significant increase in ‘disease-related activity’ was made up from various prostate-cancer related events; for example, an increased number of oncology/urology outpatient visits, diagnostic tests, treatments, admissions and SREs. Thresholds and clustering intervals were developed to identify the outcomes of interest.

In order to develop a clinically useable surrogate endpoint tool for RCTs, criteria were used for validation (section 4.9.5) (195, 196). A surrogate needed to show a strong relationship between endpoints (*patient level surrogacy*) and also show endpoint treatment effects (*trial/treatment level surrogacy*). Therefore, as undertaken in the ICECaP studies, association to existing endpoints and treatment effects were assessed.

In the early stages of training, where the clinical reference was available, the algorithm identified most reference-identified progressive outcomes, including some not documented in the trial (4.10.3). However, outcomes were also classed as erroneous if they were, 1) missed, or 2) incorrectly located either before (identified early) or after (delayed) the reference. The first type of these absent outcomes were typically due to the HES payment function resulting in omitted procedures, effecting the model.

However, outcomes were also missed if the observed number of events did not reach the threshold to detect events. Most local biochemical failures were missed due to absent routine data events. This was either due to clinical factors, resulting in an absent routine data footprint, or the limited routine data coding, both leading to activity below the threshold. For example, local prostate cancer management commonly involves the initiation of a surveillance programme or the addition or alteration of hormone therapy, neither of which are identifiable in the routine data.

Subsequent outcomes were also masked (and hence missed) due to a sustained level of activity above the threshold for several intervals. Upon analysis of these periods, the algorithm was actually found to be detecting the same progression event, with no treatment response, and so identifying when the patient had entered a continual progressive state. I therefore restricted any analysis of subsequent outcomes to the first two progressions, as subsequent clinical advantage of identifying further outcomes was deemed inconsequential, during which events became less distinct.

The second type of erroneous outcomes - those delayed or found early, occurred due to the algorithm clustering design (discussed below) and the clinical pathway for progression, reflecting the clinical difficulty in defining the events. Although, arbitrary dates are often assigned to outcomes because assigning the true date is often not possible (207).

For maximum detection of these outcomes, 8-week time-intervals were chosen to cluster the interactions. This reduced noise from non-important outcomes whilst still detecting significant outcomes. Narrower time intervals resulted in uninterpretable data noise, in contrast to wider intervals which presented a clinically unusable output (table 82). Consequently, an outcome could have occurred at any point during an 8-week period, resulting in a lower level of precision in comparison to standard trial data collection. However, when assessing the impact on the hazard ratios by grouping the STAMPEDE MFS into the intervals (figure 34, B), the results remained consistent, suggesting clustering the events had a smaller than expected clinical impact. However, due to this grouping, identifying true negatives at an individual outcome level was not possible, as the number of intervals where the patient did not progress was infinite, due

to the sustained periods of activity and delays in detection. Therefore, specificity was not calculated at an individual outcome level.

Additional algorithm rules were developed to enhance the detection of significant outcomes of interest, whilst reducing the noise from unrelated events. This involved: filtering out unrelated non-prostate cancer related interactions; prioritising events such as SREs due to the significance of these events on patient outcomes enabling the flagging of outcomes despite not reaching the threshold; including routine outpatient appointments (excluding them led to a reduction in the number of outcomes being identified); removing early confounding trial randomisation interactions and choosing the optimum threshold to detect an event. The optimum threshold that was identified, required the presence of greater than or equal five prostate cancer-related events to trigger an outcome, reducing to less than or equal four events, to enable the detection of a subsequent outcome. These rules were specific to prostate cancer and often unique to the STAMPEDE trial. However, these rules may be generalisable (or modifiable) in other oncology settings and trials.

There are also temporal issues with the coding. The algorithm was designed to identify International Statistical Classification of Diseases (ICD) 10 codes and Office of Population Censuses and Surveys Classification of Interventions and Procedures (OPCS) 4.2 to 4.8 codes. However, the coding schemes are continually updated, for example, ICD-11 and OPCS 4.9 are pending release. New classification releases can change how events are coded for. Hence, upon utilising algorithms such as those presented here, on novel data, additional codes will have to be included to capture all events of interest. This was a limitation during the development of the chapter 4 algorithm (proxy endpoint); all the relevant OPCS codes needed to be included to capture events across a wide patient timeframe but codes did change and subsequently could be used for alternative procedures. Hence, there was a potential for codes that were not relevant to have been included. This will remain a limitation where coding structures are continually updated. This may have implications on the algorithm threshold to detect events. If more prostate cancer-related interactions are being identified, the threshold may need to be raised, to reduce the number of potential false positives being detected; adaptive algorithms are required.

Another limitation of the study included the small cohort used to develop (6 patients) and train (44 patients, 41 novel) the algorithm (table 46). An overall STAMPEDE cohort of less than one hundred patients was available from the single-site QEH UHB that were linked to routine data. Hence, small subsets were only available to develop, train and then validate on this single-site cohort, prior to the validation at a multi-site level. Different trusts will code differently which has implications on the routine data. It is therefore possible that the algorithm performs differently at other sites. The differences in performance are assumed to be minimal due to the comparable hazard ratio analyses during the multi-site study (discussed below) but this cannot be confirmed in the absence of multi-site note review.

We have shown that our routine data algorithm was able to show a strong relationship between endpoints (patient level surrogacy) and endpoint treatment effects (trial level surrogacy). I found that the algorithm was not *directly* detecting a single standard trial surrogate; however, the model was detecting a combination of the standard surrogates (for example, 4.10.4.1) and the hazard ratios were comparable to those identified using the trial data (table 63).

As mentioned above, some trial endpoints are more clinically driven than others; for example, the FFS, is often clinically arbitrary, compared to the PFS. The HES endpoint was proposed to be a more clinically driven endpoint, due to identifying hospital interactions. Thus, it was hypothesised that the HES endpoint would identify outcomes more similar to the MFS and PFS. This was confirmed, the algorithm was able to identify more of the MFS and PFS endpoints, than the FFS (section 4.10.4.1). This was due to the difficulty in identifying biochemical failure, often triggering the FFS endpoint. A prostate specific antigen (PSA) rise often precedes other endpoints and is therefore a more expeditious surrogate (FFS), (151) but patients can experience an FFS event but not die from prostate cancer (9) and often these events are not clinically significant. A rise in PSA may lead to surveillance or hormone therapy change, in contrast to clinically significant events, including those that constitute the PFS, where the patient is often experiencing clinical progression. Although outcomes most similar to the clinically driven endpoints were being identified, a combination of endpoints were actually being found

and hence a novel outcome was proposed. It would be valuable to test the algorithm in another oncology setting that does not involve biochemical PSA failure, for example, breast cancer. To my knowledge there is currently no systematic way to extract UK multi-site patient laboratory results to aid the identification of FFS events. This is in contrast to Denmark, where it is possible to extract multi-site (for Northern and Central Denmark) PSA readings, hence enabling the analyses of this (173, 208). The number of STAMPEDE endpoints met could only be identified in the patients that had their clinical noting reviewed, hence the number identified were not calculated for the final large-scale validation.

My routine data surrogate was detecting the first significant increase in disease-related 'activity', thus allowing time to 'clinically relevant outcome' to be calculated in endpoint analysis – the 'activity-free survival' (AFS). This is the first most intensive period of healthcare service interaction, which I proposed to reflect the disease state of the patient. These periods are often treatment intensive and time consuming and therefore, lengthening the time to this outcome, has clear implications to patient quality of life; this correlation with quality of life is proposed to be investigated. Quality of life has been proposed to be one of the most important measures for patients with cancer (152). However, although quality of life audits for patient cohorts with cancer have been undertaken and these data are available through routine data providers (209, 210), this is not currently the standard; although there is currently a drive for routine NHS quality of life data collection. Hence, although these measures are not *directly* available in these data, an endpoint such as the AFS may provide a proxy measure for quality of life.

The correlation coefficient of the HES-derived endpoint compared to the STAMPEDE MFS and PFS (table 68), for the largest multi-site analyses, highlighted that a novel endpoint was being identified that was different to standard trial endpoints. However, there was still a strong correlation between the HES and the STAMPEDE MFS and PFS endpoint (table 68) (199, 200). In the correlation plots (figure 36), it can be seen that when compared to the standard endpoints, the HES-identified outcomes were often being found early or delayed (confirming what was identified in the early analyses). For example, more frequently, the HES endpoint was identified prior to the MFS (early). This suggested that the algorithm was more often than not, identifying progression

assessment, rather than the interventions post-progression. The correlation plot comparing the HES-detected outcomes and the trial FFS (figure 36, B), illustrated the difficulty in identifying FFS events, that was discussed above. The majority of the outcomes lay below the line of perfect correlation. Hence, HES often identified the outcomes later than the FFS, potentially highlighting that the early FFS outcomes were being missed and a later event was being identified.

The Bland-Altman analyses (figure 36) were undertaken because measures can be correlated but it does not mean they agree. The limits of agreement for the endpoints were very wide, highlighting the difference between the novel HES AFS endpoint and the standard endpoints. However, as previously highlighted no true gold standard is available to validate the events. However, Bland and Altman stated that how far apart measurements can be without causing difficulties is a question of judgement (211). Thus, although a different measure may be being identified, the utility of this novel measure remains in question until further validation has been undertaken in the trial setting.

It was not clear if a clinical pattern was being identified with the plotted outlier events, due to being outcomes that could not be validated by note review. However, the plot suggested that the HES endpoint was more comparable to the standard endpoints (MFS, PFS) for earlier outcomes (occurring within the first year of randomisation into the trial), than the later ones (figure 36). One suggestion for this could be, if a patient experiences a trial event early (near randomisation), during more frequent trial follow-up visits (for example, every six weeks), these visits may increase the total hospital interaction number above the algorithm threshold and enable earlier event detection.

Trial surrogacy was investigated; as more clinically driven endpoints are reached, such as the PFS, the effect size shown when comparing treatments often reduces. Hence, as the HES endpoint was proposed to be more clinically driven due to identifying hospital interactions, it was hypothesised that the treatment effects would be more similar to MFS and PFS, rather than the FFS. This is what was seen in the largest multi-site validation (table 66), with the HES-derived hazard ratio splitting docetaxel treatment outcomes with the standard-of-care. The HES-derived hazard ratio was closest to the

STAMPEDE MFS and the algorithm was capable of identifying treatment benefits, in line with standard endpoints. In the early stages, the confidence intervals for the hazard ratios were wide due to small patient numbers. However, upon the large-scale multi-site validation the confidence intervals became narrower, although the HES AFS comparing treatments did cross one (table 66).

Despite the efforts to create a validated endpoint with consideration to the ICECaP criteria, this study was only able to assess proof of concept and prior to consideration for acceptance as a validated endpoint within trials, much further work is required. For example, it would be necessary to perform further statistical tests of surrogacy and to compare the AFS to the OS for the same cohort of patients (151). This routine data model could also be used to compare trial to non-trial patients to test 'real-world' intervention effects and assess whether a trial result has been truly implemented in practice. However, the algorithm would also need validating within this different population.

Another thing to consider is when the results from this algorithm would be communicated to patients; this depends upon the use of the algorithm. If the algorithm was being used to capture additional events, validate standard trial collected events or be used to conduct survival analyses, then I would envisage any results to be communicated at the same time as the dissemination of the trial results. However, the algorithm could be utilised for other purposes, for example, detecting if a patient has entered a palliative care stage. I envisage the method of communicating this to the patient would be directly at the point of detecting the event (dependent on the timeliness of the routine data), as the aim would be to enhance their care.

Finally, to tie the results of this study into the current available literature, this study was assessed for weakness and bias using the developed systematic review criteria. All criteria were fulfilled (see table 81 in the appendix); however, it was not appropriate to assess the accuracy of the endpoint, compared to standard endpoints, due to the identification of a novel routine data outcome, the AFS. Although I developed the criteria, these were created to appraise the literature, prior to conducting this study, hence, were deemed to still provide a measure of weakness and bias for this study.

Further work is currently being undertaken to further validate the algorithm within different settings; for example, testing the surrogate within another drug setting (hormone therapy). Further projects are also being set up to develop the algorithm, including, for outcomes not routinely collected in a trial, for example, to assess time to subsequent treatment outcomes. Validation has also been discussed in a low-risk prostate cancer setting, an alternative prostate cancer RCT and within other oncology settings. Breast cancer would be an exciting validation, due to the disease similarities with prostate cancer (212) and radical treatment options, which are identifiable routine data healthcare interactions, as presented here. Projects are underway at the MRC CTU at UCL to conduct some of this proposed work.

4.12 Conclusion

In conclusion, routine data were again identified as having potential to improve standard clinical trial frameworks. The main aim of enhancing trial design is for patient benefit. Hence, prior to contrary belief, if routine data are used with caution, they may be a rich source of trial outcomes and may provide an alternative overall survival surrogate endpoint, for prostate cancer trials.

These two chapters have illustrated the retrospective feasibility of using routine data for trial follow-up, to identify outcomes, 1) that were *directly* identifiable within the routine data, and 2) that were not *directly* identifiable within the routine data. The next chapter describes a study undertaken to assess the feasibility of using routine data to prospectively acquire data for trial follow-up, with consideration to the evidence generated in the previous chapters.

5 CHAPTER FIVE: Direct RCT data collection (prospective follow-up model)

5.1 Disclosure

Work from this thesis chapter was submitted to the *Pilot and Feasibility* journal and is pending peer review. Work from this chapter was also presented at the *Trials using Cohorts and Routine Health Data: International symposium on their Efficiency and Analysis* conference on the 15th May 2019 and the abstract is pending publication in *Trials* journal.

5.2 Abstract

5.2.1 Background

This chapter documents the development of a novel data collection technique for trial follow-up, based upon routine National Health Service (NHS) data. Traditionally, oncology trial patient follow-up can involve patient-health care professional contact, where trial data are collected on case report forms (CRF). A novel methodology was to be developed, if deemed feasible, using data from routine administrative sources, that could provide a robust system feasible for use in a trial. To develop this, a case study was set up in the BladderPath trial. BladderPath is a bladder cancer randomised controlled trial (RCT) evaluating an alternative pathway for diagnosis and treatment, aiming to improve outcomes. This chapter reports the development and validation of a method enabling routine data use in a recruiting trial.

5.2.2 Methods

Using routine data, an algorithm was developed to identify the delay from diagnosis to treatment (number of days, time to correct treatment (TTCT)), for muscle invasive bladder cancer (MIBC) patients treated at the single site University Hospitals Birmingham Queen Elizabeth Hospital (UHB QEH). The routine data were also validated by extracting events of interest algorithmically and comparing the events to reference datasets, for example, clinical noting, to determine detection sensitivity. Subsequently, this methodology was adapted to utilise routine data in BladderPath.

5.2.3 Results

There was greater than a 3x increase in the number of days for the TTCT (68-day increase) for MIBC patients when compared to the NHS target diagnoses to treatment time target of 31 days. Overall, a total of 829/1042 events were detected using the hospital routine data. There was an increase in data quality from 2011 (41/117, 35%) to 2017 (104/109, 95%). Varying sensitivities were identified for events. A method was developed to utilise routine data using additional datasets, a querying framework and further rules.

5.2.4 Conclusion

The time from diagnosis to treatment for MIBC patients was increased compared to the NHS target, which is in line with published literature. BladderPath hypothesises that this may highlight why outcomes for MIBC are poor. The restrictive nature of the coding to identify confirmed MIBC cases, meant that events were missed. However, using routine data was shown to be a feasible method for trial follow-up. It was identified that very high sensitivities can be achieved with targeted data queries to further enhance these data. Routine data currently is not of high enough quality to solely perform follow-up without a querying framework in place. However, I believe routine data to offer a robust way of rapidly collecting trial datasets.

5.3 Introduction

5.3.1 Direct clinical trial analysis

The previous chapter (chapter 4) explored the concepts surrounding the use of routine data to perform retrospective analysis of outcomes not collected using standard clinical trial data collection techniques. The chapter found that missing, erroneous and lack of detail in these data were limitations to performing such analysis. Despite this, it was highlighted that routine data are a powerful resource for analysis for outcomes not collected within a clinical trial, but due to the limitations it was found that 'raw' (unprocessed) routine data may not be ready as a source of clinical trial data alone.

Standard clinical trial activities, such as data collection and follow-up, have been approached via face-to-face techniques (213) with patients and health care professionals. Trial data can be collected manually on case report forms (CRF) during patient follow-up visits (214) but substantial time is often required for this (215). Thus, moving to using a routine data framework would be a radical departure from usual methods.

This chapter outlines the feasibility for the design and implementation of a novel clinical trial methodology using routine data for patient follow-up. If the methodology was deemed feasible to provide a robust system for a trial, the next aim was to implement a clinically useable method for an RCT, with BladderPath as the example. BladderPath is funded by NIHR HTA (National Institute for Health Research, Health Technology Assessment) (see introduction, 1.5.2). The overall aim of BladderPath is to conduct an RCT achieving broad patient recruitment with the least clinical, cost and resource disruptions.

As mentioned in the introduction, the TTCT for all patients (both non-muscle invasive bladder cancer (NMIBC) and MIBC) are collected as a primary outcome for the BladderPath trial (see introduction 1.5.4) (73). This is to compare the standard treatment pathways with the split enhanced pathway. Overall survival and recurrence are also to be collected in the final clinical stage, to assess if a reduced time to correct treatment (TTCT) does indeed improve outcomes in the MIBC patient subset (73).

5.3.2 Chapter rationale

Using the BladderPath trial as an example, in this chapter, I report the development and validation of a novel methodology which I developed to enable routine data use for trial follow-up.

5.4 Background

5.4.1 The dawn of routine data in UK RCTs

The idea of using routine data to enhance RCTs is not new. In 2000, Lewsey (11) published a report performing studies to explore how routinely collected data could be used to '*complement or supplement RCTs*' (11). Scottish morbidity records were analysed for hospital episodes and death (1981 to 1995) but these data were not linked to an RCT. The aim of the individual studies were to answer questions that would not be appropriate to perform using standard RCT designs (as in the chapter 3 study). For example, the procedure would be considered unethical through randomisation, or the sample size would not have been large enough through traditionally recruited studies (11).

The study supported the conclusion that '*routinely assembled NHS data might have value as a complement or alternative to RCTs in certain circumstances*' (11). However, the authors noted limitations to conducting such studies, concluding, '*studies based on routine data fall short of the rigour that one expects in RCT designs when comparability between the different arms of a trial is fundamental to the assessments that are made*' (11). At this time, the scope of routine data was identified but the quality made it not feasible to consider using these data in an RCT.

In 2003, Williams reported that '*routinely captured clinical data have real potential to measure patient outcomes, particularly if the detail and precision of the[se] data could be improved*' (216). But it was not until 2005, that the Health Technology Assessment (HTA) then recognised that there was a vicious circle, '*as the[se] data are poor, they cannot be used; lack of use ensures they remain limited and of poor quality*' (217). Hence, the HTA provided proposals to enhance data quality with the aim to make these data a useable resource for RCTs (217).

5.4.2 Validating UK RCT routine data

Since 2005, many studies have linked routine data to RCT data, however, not for trial conduct but for validation purposes (46, 103, 105). For example, Wright-Hughes investigated the use of routine Hospital Episode Statistics (HES) data for identifying self-

harm outcomes for trials (105). This was a validation study as data were first collected via standard researcher visits to site. They concluded that '*further targeted data collection through researcher site visits*' was required for supplementary information (105). Kilburn (103) explored the use of routine data by linking trial data from an oncology trial (TACT) to routine data, however, did not use routine data for follow-up. The four breast cancer trials under investigation within this study are, TACT2, POETIC, IMPORT HIGH and FAST FORWARD, which are completed trials; analysis is also being undertaken into identifying endpoints in the routine data (218). Powell also validated the use of routine data in an epilepsy RCT setting (219).

Routine data have also been used frequently for epidemiological outcome generation. For example, in prostate cancer, the trial PHONIC, used routine data to determine rates of osteoporotic fracture, but routine data were not used for patient follow-up (102). The Cluster Randomized Trial of PSA Testing for Prostate Cancer (CaP) team also performed retrospective linkage for a health economics study (220).

All of these studies performed retrospective analysis of routine data linked to trials. The aims were either to identify an outcome, or for validation purposes: no study used routine data to provide an alternative method of RCT data collection. However, there has been an increase in the number of RCTs using routine data to supplement or replace traditional trial data collection and is discussed below (5.4.3).

5.4.3 Routine data supplementing follow-up in UK RCTs

Trials have been using routine data to supplement standard outcomes collection. For example, mortality records are commonly accessed to supplement trial data collection (156). The Knee Arthroplasty Trial (KAT) (221) used routine English HES data in addition to the Scottish equivalent (ISD) to supplement the trial data (221). Admissions and other surgical events were identified in the routine data, upon which, the clinical noting was queried (221). Trials have also used routine data for long-term follow-up of outcomes. The REACT (Randomised Evaluations of Accepted Choices in Treatment) trials were developed to use General Practice Research Database (GPRD) primary care data (2010). Patients were monitored as usual in clinical practice, then the general practice (GP) electronic health record from each visit were downloaded daily into the trial database

for side effect monitoring. The trial database was then compared to the GPRD database and the routine data used for long-term follow-up (222). The two REACT trials detailed were Retropro (a cardiovascular study, ISRCTN33113202) and eLung (a chronic obstructive pulmonary disease study, ISRCTN72035428) (222-224).

In an oncology setting; Bhattacharya is investigating the use of routine data to supplement the recruiting breast cancer study: PRIMETIME (218). They aim to download NCRAS routine data at six-monthly intervals for follow-up. However, PRIMETIME also uses traditional patient follow-up for ten years, so is not using routine data exclusively (218, 225). Fitzpatrick (35) also investigated the use of routine data for long-term RCT follow-up. Fitzpatrick found 113 trials worldwide had been extended by record linkage, including 25 in the UK, concluding that routine data linkage is an underused approach which may add insight compared to traditional data collection methods (35).

5.4.4 Routine data as the basis for follow-up in UK RCTs

Two cluster primary care RCTs were designed to utilise routine data for follow-up, one for antibiotic prescribing for acute infection (226) and the other looking at prevention after first stroke (227). It was identified that both RCTs could be performed efficiently using electronic healthcare records (CPRD) and were identified as being among the first cluster RCTs to be '*performed exclusively*' using electronic health records. The electronic health records were populated from the information recorded by the clinician during the visits (228). No secondary care RCTs or oncology RCTs were identified to use routine data as the basis of data collection.

In summary, the scope for routine data use in RCTs has been recognised for two decades (11), however, until recently RCTs only used these data for validation (103) or outcome generation (102). More recently, routine data have been used to supplement RCT data collection (221). However, no secondary care RCTs have been identified to use routine data sources as the basis for patient follow-up, including none identified in oncology. A method was therefore to be designed, if feasible, to utilise routine data as the method of patient follow up for a secondary care oncology RCT, the example being BladderPath.

5.5 Materials and methods

5.5.1 Summary methods

5.5.1.1 Approvals

All data were sought with relevant ethical and regulatory processes in place (section 2.4). Both the time to correct treatment analysis and the data validation were registered as audits at UHB QEH. Routine Hospital Interactions Data (HID) (a HES data equivalent) were provided by the informatics department at UHB QEH. The audits conducted for this validation study are also referenced in the BladderPath trial protocol (see section 2.4 for further details) (73).

5.5.1.2 Summary participants

Two different analyses were undertaken, time to correct treatment (TTCT) and data quality analyses. All participants included were patients that underwent cystectomy surgery (bladder removal) to the bladder (not exclusively for bladder cancer) or bladder radiotherapy (radical or palliative) at UHB QEH. This included 206 patients that had bladder cystectomy surgery (not exclusively for bladder cancer) between 08/01/10 – 07/04/17 identified in the manual surgical reference group and 525 patients identified from a radiotherapy reference (radical and palliative), that were treated between 01/01/11 – 11/06/18. The surgical reference was designed to collect the desired cohort and therefore did not require further cohort extraction. However, the radiotherapy cohort were extracted by using the ICD (32) bladder cancer code C67X.

The data quality validation cohort was not restricted by bladder cancer and as such, the whole cohort of unique patients were eligible for analyses. In contrast, the patient analysis profile for the TTCT can be seen in figure 39, where a diagnosis of MIBC was required.

The programming language SQL was used to extract the patient characteristic information from the hospital interactions/informatics routine data (note that these data were extracted by A. Dosanjh, in table 72, for the data quality analysis cohorts only). This included counts and percentages for gender, ethnicity and Charlson

Comorbidity Index scores ('Charlson scores') (143). The Charlson scores were calculated if the patient had experienced an inpatient procedure. The score was calculated at the time of the surgery to the bladder or if the patient had not had surgery, the nearest inpatient admission to the start date of the radiotherapy was identified. In addition, the median, Interquartile range (IQR) and range was identified for age. For the data quality analyses, age was chosen as, the age at first treatment identified in the routine data and in the TTCT analysis, the age was identified as the age at flexible cystoscopy (a proxy for diagnosis) identified in the routine data. I calculated the patient characteristics for the TTCT analyses manually from the clinical noting and therefore Charlson scores were not identifiable (table 72).

5.5.1.3 Data sources

The data source used in both analyses were the local routine HID. As these data are returned centrally to National Health Service Digital (NHSD) to form the HES, they were assumed to be an equivalent to the HES data (2). No data linkage was required as linkage to a clinical trial number was not necessary. Comparison to reference data sets was possible via unique hospital identifiers.

5.5.2 Joint methods

The algorithms to extract events of interest from the routine HID were written and run in R (109) using RStudio (110) for both the TTCT analyses and the data quality validation (code available on request).

5.5.3 Time to correct treatment (TTCT) analysis

5.5.3.1 Participants

The participants investigated in the TTCT analyses can be seen in figure 39.

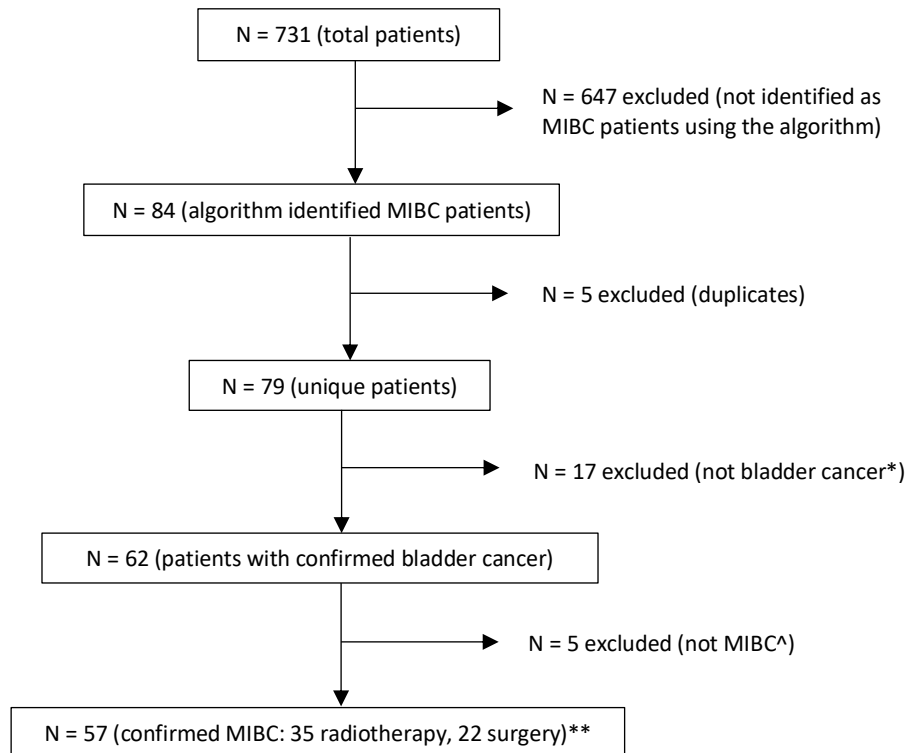


Figure 39: The study profile for the cohorts utilised in the TTCT analyses.

* = 17 patients were excluded when the routine data-identified diagnosis dates (flexible cystoscopy) were queried with the clinical noting; these patients were suspected to not have bladder cancer. The 17 exclusions included: The upfront diagnosis was missed (N=9); flexible cystoscopy was used for a previous renal condition (N=4); incorrect routine data (N=2) or bladder cancer was not yet formally diagnosed (N=2). ** = The final cohort consisted of 35 patients being given radiotherapy and 22 surgery for MIBC (patients without bladder cancer were excluded using the algorithm and note review). ^ = Also identified upon note review.

5.5.3.2 Analytical methods

The algorithm was developed to apply proxy rules to the routine HID to extract patients with suspected *de novo* (newly diagnosed) MIBC following the standard pathway. The time to the following events (in days) was calculated for:

- 1) Diagnosis (flexible cystoscopy) to transurethral resection of bladder tumour (TURBT) (defined by the NHS as the first definitive treatment for bladder cancer) (see introduction figure 4; time X),
- 2) TURBT to correct treatment (see introduction figure 4; Y) and

- 3) Overall time from diagnosis to treatment (see introduction figure 4; X+Y).

The codes to identify events from the routine data can be seen in section 8.3.1 (appendix, table 91). These time to event values were calculated using R, and the mean, minimum and maximum values calculated using Excel. The difference in the time to the correct treatment of surgery and radiotherapy were also compared.

For each patient, each of the routine data identified flexible cystoscopies were confirmed via note review to confirm a bladder cancer diagnosis. If the flexible cystoscopy was not performed to identify *de novo* bladder cancer, the patient was excluded from the analysis. A random sample of 10 patients' pathways (30 events, 10 flexible cystoscopies, 10 TURBTs, 10 subsequent treatments) were confirmed against the clinical noting to confirm the accuracy of the routine data in detecting individual events. All patients were assessed in the clinical noting for MIBC diagnosis to calculate sensitivity and positive predictive value (PPV) accuracy of MIBC detection using the proxy rules (table 69).

Rule/ process	Description
1	Extract the first flexible cystoscopy event for each patient
2	Extract the first TURBT event for each patient
3	Keep only the events where patients have 1) both a flexible cystoscopy and a TURBT event, 2) the flexible cystoscopy must come before the TURBT
4	Extract the first definitive treatment event (radiotherapy, surgery, chemotherapy) per patient
5	Keep only the events where patients have had 1) a flexible cystoscopy, a TURBT event and definitive treatment, 2) the flexible cystoscopy must come before the TURBT which both must come before the definitive treatment
6	Time to event calculations: Time to the TURBT minus flexible cystoscopy, definitive treatment minus TURBT, definitive treatment minus flexible cystoscopy

Table 69: The algorithm process to extract suspected MIBC patients to enable time to definitive treatment calculation.

5.5.4 Data quality assessment

5.5.4.1 Data sources

Four data sources were used to analyse the surgical and radiotherapy cohorts:

- 1) Radiotherapy data: Consisting of radiotherapy machine prescription data.
- 2) Manually collected surgical data: Used to identify the surgical cohort. This is a dataset maintained by the surgical team to collect manual data on all surgical patients (for example, from data extracted from the clinical noting). In addition, the manually collected data were enhanced with data from the national British Association of Urological Surgeons (BAUS) dataset (229, 230).
- 3) The routine HID (HES-equivalent) with inpatient and outpatient interactions (206). These data were extracted by the informatics team at UHB QEH using the NHS number and the hospital number which were both identified in the reference cohorts. In order to capture events prior to the manually documented date of cystectomy, the routine HID for the cystectomy cohort were extracted one year before the documented reference date of surgery. These data were censored at 31/03/18. To extract the routine HID for the radiotherapy cohort, data were extracted from the first radiotherapy event that was documented in the radiotherapy reference which was the 01/01/11. These data were censored at 31/05/18.
- 4) UHB QEH clinical note review data, for example, correspondence and clinical noting.

Reference data were assumed to be the 'gold standard' to compare to the routine data to enable accuracy of detection to be calculated. Reference data consisted of three sources: 1) the manually collected surgical data, containing procedure and date, used to validate the surgical routine HID events, 2) the radiotherapy administrative data, used to validate the radiotherapy routine HID events, 3) the note review extracted data from the clinical portal, used to validate the following routine HID events: Chemotherapy, Bacillus Calmette-Guérin (BCG), cystoscopy, and censor (date of last known clinical visit – a follow-up trigger) (table 70).

Test data were the routine data that were being validated for accuracy against the reference and consisted of one source, the routine HID (table 70).

Data source	Data type	
	Reference	Test
Radiotherapy administrative data	✓	✗
Surgical data (manually collected)	✓	✗
Routine HID (inpatient and outpatient interactions)	✗	✓
Clinical note review data	✓	✗
BAUS national dataset (surgical data enhancer)	✓	✗

Table 70: *The data sources utilised for the validation study.*

5.5.4.2 Participants and events of interest

For the data quality assessment, a total of 707 patients had at least one event of interest validated across the surgical and radiotherapy cohorts. For surgical event accuracy, the whole cohort of 206 patients was used, and random subsets of events that were identified in the routine HID were used to check other events. The number of events by participants that were validated can be seen in table 71. In addition, the whole radiotherapy cohort of 525 patients were used to validate the radiotherapy events (table 71).

Event	Number of patients	Number of regimens validated	Number of individual events (procedure or administrations) validated
Surgery to bladder	206	NA	206
Chemotherapy	40	47	-
Cystoscopy	29	NA	106
BCG	30	15	114
Censor date	100	NA	100
Radiotherapy	525	568	7894*

Table 71: *The number of patients, number of regimens and number of individual events (administrations, radiotherapy fractions), validated by event type.*

*Patients were able to have multiple regimens and each regimen could include multiple administrations (for example, BCG and radiotherapy but individual administrations were not assessed for chemotherapy); * = the number of radiotherapy fractions in the 568 regimens; NA = not appropriate; '-' = not calculated.*

5.5.4.3 Processing and outcome measures

To identify the events of interest from the routine data (seen in table 71), the algorithm was designed to extract coded routine data events (Office of Population Censuses and

Surveys Classification of Interventions and Procedures (OPCS) codes version 4.4 – 4.8) (33). In addition, the censor date was extracted from the NHSD main speciality (MAINSPEF) coding in the informatics data (231) (see appendix 8.3.2 for the codes that were extracted, appendix table 92). The events of interest were:

- Surgery to the bladder to remove the tumour: Including cystectomy, cystoprostatectomy and exenteration procedures.
- Radiotherapy - Including radical (for curative effect) and palliative (for pain relief) regimens.
- Cystoscopy: Including all types of cystoscopy, including but not limited to flexible procedures (cystoscopy or urethroscopy) or rigid TURBT (therefore events could be both before or after cystectomy).
- BCG immunotherapy therapy.
- Chemotherapy: Including treatment for any cancer, as it is not possible to identify what the chemotherapy was administered for in the outpatient routine HID.
- The last known interaction that the patient had, either with urology or oncology departments. These events were identified as inpatient or outpatient events.

5.5.4.4 Analytical methods

The events that were identified using the algorithm were manually compared, by date, to the reference events to calculate the PPV and sensitivity. Events were also grouped by year to assess if the sensitivity changed over time. Assessments were undertaken where, 1) the event was required to be detected on the exact reference date, 2) the event was not required to be identified on the exact date. The exact diagnostic code was not required to detect the events, but the correct type of code was required. For example, if the wrong banding of chemotherapy was identified in the routine data, the chemotherapy event was still classified as being identified. This was due to the restrictive nature of the coding identified in chapter 3 and 4.

For outpatient events, for example, flexible cystoscopy, the date of appointment was validated against the reference data. Operation dates were validated for inpatient surgical procedures. Regimen level accuracy was investigated for BCG, radiotherapy and

chemotherapy events, only one event per regimen was therefore required to be present in the routine data to correctly identify that the regimen had been administered.

5.5.4.5 Follow-up design outcomes

Using the results from the studies, if deemed feasible, a framework was to be designed for use in RCTs, using the example, the BladderPath trial.

5.6 Results

5.6.1 Participants

The patient characteristics for all cohorts analysed in this chapter can be seen below in table 72.

		TTCT analysis	Data quality analysis cohorts	
		All cohort	All radiotherapy	Surgical
		N=57	N=525	N=206
		Number of patients (%)		
Age at 1st treatment	Median (IQR)	72 (64-79)	75 (68-94)	66.5 (56-73)
	Range	50-94	31-96	22-85
Gender	Male	36 (63%)	380 (72%)	147 (71%)
	Female	21 (37%)	144 (27%)	59 (29%)
Ethnicity	White	47 (83%)	401 (76%)	191 (93%)
	Asian/Asian British	2 (4%)	18 (3%)	9 (4%)
	Black/Black British	2 (4%)	5 (1%)	3 (2%)
	Mixed	2 (4%)	7 (1%)	0 (0%)
	Other	1 (2%)	2 (0%)	2 (1%)
	Unknown	3 (5%)	91 (17%)	1 (1%)
Charlson score	<1	-	253 (48%)	116 (56%)
	1-5	-	73 (14%)	45 (22%)
	6-10	-	39 (7%)	23 (11%)
	11-15	-	21 (4%)	10 (5%)
	16-20	-	11 (2%)	5 (2%)
	>20	-	8 (2%)	3 (2%)
	Unknown	-	119 (23%)	4 (2%)

Table 72: Patient characteristics for the TTCT assessment and the data quality analyses for the cohorts.

In the data quality analyses, the patients without an inpatient event in the routine data have an unknown Charlson score. (-) = The Charlson scores could not be identified from the clinical noting for the TTCT analyses and hence are not presented in the table. The percentages in the table may not sum to 100% due to the rounding.

In the total surgical cohort (N=206), 106/206 (51%) patients were diagnosed with MIBC, 64/206 (31%) had NMIBC and the remaining 36/206 (17%) did not have cystectomy for bladder cancer. All patients in the radiotherapy data were extracted upon the condition of a bladder cancer diagnosis, however, the proportion of MIBC to NMIBC was not identifiable.

5.6.2 Time to correct treatment (TTCT) analysis

29/29 (100%) events (flexible cystoscopy, TURBT and subsequent correct treatment) that were queried to the reference clinical noting were correctly identified in the routine data (one event could not be confirmed). The routine data identified one event one-day early, when compared to the reference (table 73).

Validated event	Reference events	Routine data identified
Flexible cystoscopy	10	10
TURBT	10	10
Correct MIBC treatment	9	9
Total	29	29

Table 73: Validating the routine data algorithm coding to identify the correct events in a small cohort.

The algorithm proxy rules to identify MIBC were investigated. Compared to the surgical cohort, where it was known that 106/206 had MIBC, only 35/106 (sensitivity: 0.33) patients were identified to have MIBC using the algorithm rules using the direct coding.

Across both cohorts (surgical: 35, radiotherapy: 44), the algorithm identified 79 unique patients (figure 39) as having MIBC, two of these had obvious inaccurate routine data outcomes, so the accuracy of detection before exclusions was calculated for 77 patients. 57/77 routine data events were true MIBC events and there were 20 false positives (PPV: 0.74) (figure 40, A). Hence, before exclusion of non-bladder cancer patients using the reference, many events were being missed and many false positives were present. After non-bladder cancer patients were excluded using the reference, 62 patients were identified with the proxy. 57/62 were confirmed as MIBC events (PPV: 0.92), hence restricting the cohort increased the PPV (figure 40, B).

A		MIBC reference		
		Yes	No	
Algorithm proxy test	Yes	57	20	77
	No	NA	NA	NA
		57	20	

74.0

B		MIBC reference		
		Yes	No	
Algorithm proxy test	Yes	57	5	62
	No	NA	NA	NA
		57	5	

91.9

Figure 40: The PPV for identifying the MIBC patients.

A) The routine data extracted events prior to excluding non-bladder cancer patients using the reference (bladder cancer and non-bladder cancer patients present); B) after excluding non-bladder cancer patients (only bladder cancer patients present).

Using confirmed MIBC patients, the routine data acquired time from diagnosis to starting TURBT was 27 days but for these MIBC patients where further correct treatment was required, there was over a 3x increase (99 days in total) in the number of days to treatment than the 31-day target. Post-TURBT it was quicker to have surgery than radiotherapy, with a mean difference of 11 days (radiotherapy = 75 days, surgery = 64 days). 71 further days were required from TURBT until correct treatment was given (surgery or radiotherapy) (figure 41).

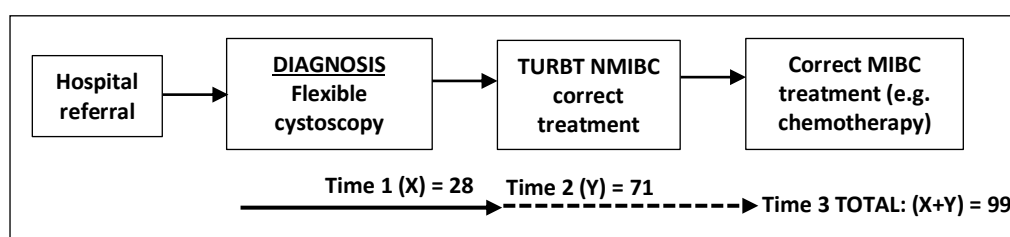


Figure 41: The time to definitive treatment for MIBC patients.

X = time from diagnosis to correct treatment for NMIBC; Y = further time to correct treatment for MIBC; X + Y = total time from diagnosis to treatment for MIBC.

5.6.3 Data quality assessment

In total, the algorithm utilising the routine data detected 829/1042 events (sensitivity: 0.80) across the ten data years (2008-2018) (table 74).

Event	Total events [reference]	Total events [routine data]	Accuracy (%)
Cystectomy	206	202	98.1
Radiotherapy regimen	568	391	68.8
Censor	100	89	89.0
BCG regimen	15	14	93.3
Cystoscopy	106	89	84.0
Chemotherapy regimen	47	44	93.6
Total events detected	1042	829	76.9

Table 74: The total number of events that occurred (identified in the reference), compared to the total number of events that the routine data were able to identify.

A variation of this table is in a publication in submission.

The number of events split by year from 2011 can be seen in table 75. The data quality improved 60.4% from 2011 to 2017 (the years where greater than 100 events were identified in the reference) (table 75).

Year	Total events [reference]	Total events [routine data]	Sensitivity %
2011	117	41	35.0
2012	144	56	38.9
2013	139	118	84.9
2014	146	142	97.3
2015	149	145	97.3
2016	149	148	99.3
2017	109	104	95.4

Table 75: The number of events identified in the routine data and compared to the reference (2011-2017). The sensitivity is displayed as a percentage (%).

Over the last five full data years (2013-2017) years, the routine data had a sensitivity of 95% (657/692) (figure 42).

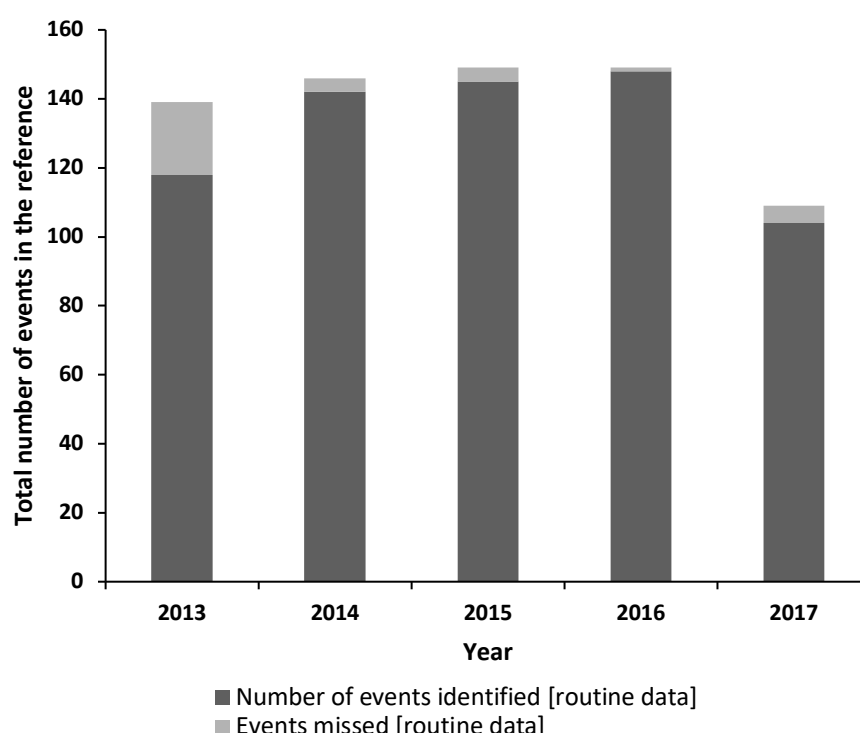


Figure 42: The number of events identified in the last five full data years (2013-2017).

In the surgical cohort, a 100% linkage rate of the reference data to the routine HID was present, 206/206 patients had a routine HID event documented (any inpatient or outpatient interaction, not exclusive to those analysed). In the surgical cohort, using the whole ten-year data period, when the exact date of surgery to bladder was required to correctly identify an event, the sensitivity was calculated as 96%; with 198/206 events correctly identified. When not restricting the routine data event detection to a date, a sensitivity of 98% was present with 202/206 surgical events identified (table 74). These four additional events were detected less than two weeks from the date of procedure as identified in the reference. Two 1-day, one 4-day and one 13-day delay in event detection were identified. In addition to this, a PPV of 96% was calculated, as eight false positives were also detected. This was due to duplicate events, unrelated events (for example, TURBT) and procedures that were not undertaken. The quality of the routine data was consistently high across all data years.

Chemotherapy regimens could be identified with a sensitivity of 94% (44/47 events were detected) (table 74). The PPV was 94% for detecting chemotherapy regimens, the three false positive regimens that were identified were due to wrongly detecting BCG

treatment (table 76). 100% of chemotherapy regimens were identified from 2012 onwards.

A sensitivity of 84% was present for detecting cystoscopy events (89/106 were detected) (table 74). 100% of TURBTs were identified (32/32) and 77% of flexible cystoscopy events (41/53 were detected). The PPV for detecting cystoscopy events was 94% (table 76). The six false positives were due to: duplicate records (three events), an extirpation of bladder lesion (one event), a nephrostogram plus insertion of stent (one event) and cystodiathermy (one event). Cystoscopies are also often undertaken during the three false positive procedures identified, hence a potential reason for the detection of these events.

At a regimen level for BCG (only one administration was required to be identified), 14/15 regimens were identified (sensitivity: 0.93) (table 74). In contrast, when assessing the routine data accuracy to identify all individual administrations in each regimen, 114/149 administrations were identified (sensitivity: 0.77). The majority of the false positive regimens that were detected (20 regimens), which reduced the PPV to 41%, were due to Mitomycin C administration (due to a similar method of administration) (table 76).

Finally, the only event to show a reduction in data quality over two consecutive years post-2014, was the censor follow-up event. A total sensitivity of 89% was present (89/100 events were detected) (table 74).

A 100% linkage rate was not present for the radiotherapy cohort. Of the 525 patients in the reference, only 524 patients had an event in the routine data. Only 69% of regimens were identified (391/568) (table 74). The PPV was high at 95%, with only 20 false positive regimens being detected (table 76). The data quality improved from 2011 to 2017. In 2011 a sensitivity of 1.4% was present (1/74 events detected) to 100% in 2017 (68/68 events detected) (figure 43). 7894 individual fractions were administered across the 568 regimens; of these, 5121 were identified in the routine data (sensitivity: 0.65).

	Number of events	Year of event											False positives (PPV)
		2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	
Cystectomy	Reference	NA	NA	16	21	28	26	41	34	34	6	NA	8 (0.96)
	Routine data	NA	NA	16	21	25	25	41	34	34	6	NA	
	Sensitivity (%)	-	-	100.0	100.0	89.3	96.2	100.0	100.0	100.0	100.0	-	
Radiotherapy regimen	Reference	NA	NA	NA	74	83	72	67	79	93	68	32	20 (0.95)
	Routine data	NA	NA	NA	1	1	55	66	79	92	68	29	
	Sensitivity (%)	-	-	-	1.4	1.2	76.4	98.5	100.0	98.9	100.0	90.6	
Censor	Reference	NA	NA	1	5	4	8	16	11	9	27	19	0 (1.00)
	Routine data	NA	NA	1	5	4	8	16	11	9	22	13	
	Sensitivity (%)	-	-	NA	100.0	100.0	100.0	100.0	100.0	100.0	81.5	68.4	
BCG regimen	Reference	NA	NA	1	3	1	4	2	3	1	NA	NA	20 (0.41)
	Routine data	NA	NA	1	3	0	4	2	3	1	NA	NA	
	Sensitivity (%)	-	-	NA	100.0	-	100.0	100.0	100.0	-	-	-	
Cystoscopy	Reference	NA	3	8	8	21	20	15	15	9	5	2	6 (0.94)
	Routine data	NA	2	6	6	19	17	12	11	9	5	2	
	Sensitivity (%)	-	66.7	75.0	75.0	90.5	85.0	80.0	73.3	100.0	100.0	100.0	
Chemotherapy regimen	Reference	1	1	4	6	7	9	5	7	3	3	1	3 (0.94)
	Routine data	0	0	4	5	7	9	5	7	3	3	1	
	Sensitivity (%)	-	-	100.0	83.3	100.0	100.0	100.0	100.0	100.0	100.0	-	

Table 76: The sensitivity and PPV estimates of the routine data (2008-2018) coding when events were compared to the reference.

NA = no events were validated (due to the random sample selected, or due to the data censor, for example, surgery to bladder censor mid-2017). (-): if only one event, or none were validated, the sensitivity was not calculated due to the sample size. This table is currently in the process of being published in a paper in submission.

The accuracy of the routine data algorithm to detect the correct MIBC treatments (surgery, radiotherapy and chemotherapy) can be seen in figure 43. As previously stated, the TTCT is a BladderPath primary outcome measure.

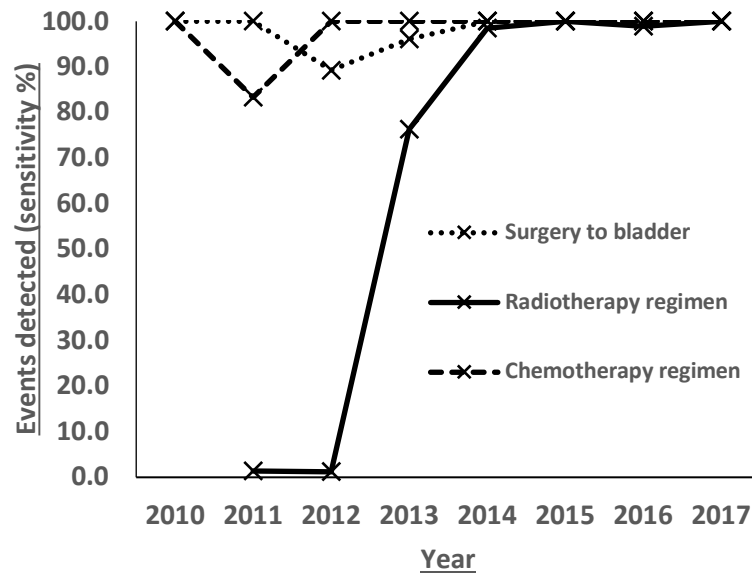


Figure 43: Accuracy of the routine data to detect the BladderPath primary outcome measure.

5.6.4 The framework

Due to the false positive and false negative events identified during these studies, plus the knowledge gained during chapter 3 and 4, I believed *directly* using coding to identify events for a trial using the hospital administration data alone, was not feasible.

However, I propose three techniques may enable the feasibility of this, 1) additional routine datasets, 2) a querying framework and 3) additional rules.

The overall proposed schema designed to enable the feasibility of using routine data as the basis for trial follow-up, can be seen in figure 44. An algorithm to extract events, such as the one presented here, could be run down the routine data at frequent intervals, to identify the events of interest. These events could then be incorporated into a CRF which could be electronically verified at the sites using queries to either modify, confirm or reject the event before being uploaded to the trial database.

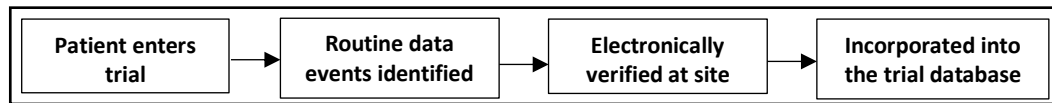


Figure 44: *The methodology developed for the BladderPath trial*

The results presented within this chapter suggest that routine data may not be of high enough quality for a trial without implementing additional measures. The results are discussed in section 5.7.

5.7 Discussion

As identified in the literature, although clinical trials have used routine data to supplement standard clinical trial data collection for many years, there is limited evidence of studies using routine data to replace these standard follow-up techniques. Standard techniques may include patients visiting a trial clinic for consultation and CRF completion to document any trial events. This method of data collection is resource and time intensive and of financial burden. In this thesis I assessed the feasibility of using routine data to strive to provide an alternative methodology to alleviate these complications.

5.7.1 Time to correct treatment (TTCT) study

The time from diagnosis of bladder cancer to the TURBT (the correct treatment for NMIBC but incorrect treatment for MIBC) was three days less than the published NHS target value of 31 days. This suggested that although the sample of patients analysed in this TTCT study were a MIBC cancer cohort, the NMIBC would be treated within the target. However, for the MIBC patient cohort, where further treatment is required post-TURBT, there was over a 3x increase in the number of days to treatment (with an additional 71 days from TURBT to correct treatment (surgery or radiotherapy), when compared to the 31-day target. The delay in time to treatment for MIBC were comparable to other studies across many different countries (225), giving confidence that the results identified by the routine data are representative of the population. It was also identified that there was a mean 11 day decrease in the TTCT if surgery was given, when compared to radiotherapy.

A limitation of this TTCT study is that only single-site data (from UHB QEH) were used. This meant that the routine data could not be validated across multiple sites nor could the delay to treatment be assessed in more depth. As discussed previously, different sites will have differences in their coding but also there will be differences in the clinical pathway, impacting waiting times. Hence, multi-site analyses would allow comparisons.

To confirm patient inclusion in the time to event analyses, the routine data detected flexible cystoscopy events per patient were confirmed by note review to include only

patients with MIBC. In the majority of the excluded cases, the routine data did not identify the initial flexible cystoscopy diagnosis due to the timeliness of the routine data provided. For example, when diagnosis occurred many years previously and the surveillance cystoscopy was identified as the event of interest. Patients were also excluded where the routine data identified exploratory diagnostic events, led to an uncertain or non-bladder cancer diagnosis. In this study, the ability to confirm events using the note review enabled these false positive MIBC events to be excluded from the sample. This enhanced the integrity of these data in this case. However, in the absence of a reference, these events would have been erroneously included in the analyses. Therefore, it was necessary to validate these routine data events prior to inclusion in the analyses. When only MIBC patients were included in the analyses, by restricting the cohort during the note review, the number of false positives reduced and there was therefore a large increase in PPV (figure 40).

As seen in previous chapters, it was hypothesised that MIBC patients were being falsely excluded (missed) from the analysis due to missing/erroneously coded routine data events. This can be illustrated by the number of patients in the surgical cohort that were identified as having MIBC using the algorithm proxy routine data rules, when compared to a MIBC cohort reference. In the reference, 206 patients underwent cystectomy, 170/206 patients had a radical cystectomy for bladder cancer and 106/170 (62%) of these patients had MIBC. However, the algorithm proxy only identified 35/106 (33%) of eligible MIBC cases.

These missing patients are hypothesised to be due to the, 1) routine data coding, or 2) patients following a different treatment pathway. Due to routine data coding, the extraction of single codes to detect individual bladder cancer events restricted detection. As also identified in the chapter 3 analyses, broader coding needed to be algorithmically extracted to increase the sensitivity of event detection; however, this technique would reduce the specificity. MIBC patients may have also been missed that did not follow the same treatment pathway as that used to extract the patient cohort, due to clinician preference or patient illness. Although not present in this cohort, due to all patients being known to receive a correct treatment (surgery or radiotherapy), it is possible that when using this model more widely this would exclude patients that never

received the correct treatment, for example, due to illness. As time goes on, the implications of the BladderPath trial (replacing TURBT with MRI imaging for MIBC patients) could also change the pathway and therefore not enable detection of these patients using the model developed here. Hence, in order to capture more patients, additional pathway rules should be added to the algorithm to aim to capture further events (an adaptive design is required). A sample of the 71 MIBC patients that had a cystectomy, that were missed from the analyses, could be studied to identify why they were missed; the routine data should be compared to the clinical noting.

A small sample of all events of interest were also validated to confirm the accuracy of the routine data events being detected. For example, that the OPCS codes M459 (unspecified diagnostic endoscopic examination of bladder) and M421 (endoscopic resection of lesion of bladder) were identifying flexible cystoscopy and TURBT respectively. 100% of flexible cystoscopy, TURBT, and subsequent correct treatment (surgery, radiotherapy) events were identified (table 73).

To reduce the problems identified with missing (reducing the sensitivity) and false positive data (reducing the positive predictive value), it was hypothesised that increasing the number of included codes to identify the outcome could increase the number of events identified. However, an additional data query technique would be required to increase data quality to reduce the number of false positives. This method of data query was tested when querying the flexible cystoscopies above. The benefit of increased accuracy of these data outweighed the burden of the querying. This burden was low due to the ability of the routine data to pinpoint individual events with a time stamp, enabling *direct* identification by date and procedure in the clinical noting. This again suggested that due to routine data quality limitations, a querying mechanism would be required to enable the feasibility of routine data in trials where routine data are the primary data source. The quality of the routine data was assessed further during the data quality validation study.

5.7.2 Data quality validation study

The Hospital Interactions Data (HID), the HES-equivalent, was seen to be a suitable dataset, in the absence of HES data. The quality of the HES data, if anything, would be

higher due to the further processing of this SUS data prior to becoming HES (206) (figure 2). Therefore, it was deemed a suitable alternative for the analyses.

As seen in table 76, events were found to be missing but the proportion of missing events reduced over time. In fact, the quality of all of these data items except the date of the last visit (censor) reached 100% in the last year available for analysis (2017) (table 76). The radiotherapy events historically had the lowest coding accuracy, however, this accuracy increased in 2013/2014 and was sustained, although small fluctuations were present (figure 43). I investigated this increase in accuracy; anecdotally it occurred due to a coding review at the trust. During this reporting period, increased checking procedures were implemented to ensure correctly coded events. Due to the impact on remuneration to the hospital (232), the accuracy has remained consistently elevated since. In parallel to this, an NHS audit was undertaken in 2013/2014, across 50 trusts. This audit was specifically looking into the coding of co-morbidities, as only those relevant to the period of care are required to be documented. It also identified a trust with a maximum primary procedure code error rate of 26%. A report was released following these audits, including, a ten-step checklist to enhance data quality. The function of reports are to increase data quality, the results of which could be seen during the data validation study (233). Due to the primary payment function of these data (232) and central initiatives to increase data quality (233), I hypothesised that recent data across other sites (not just limited to the UHB QEH site) would also reflect this pattern with more recent data being more accurate. However, as discussed above for the TTCT study, the validation of only single-site data did not enable the routine data quality to be analysed across sites which is a limitation. Another limitation in the TTCT and data quality study was that neither the specificity nor the negative predictive values could be calculated; I did not have access to a reference documenting the patients that did not have a particular event.

Missing data had implications on the Charlson scores, as some were scores missing. This is because the inpatient admission events often occur at different sites to where radiotherapy was administered and therefore would not be identifiable in the local routine data. Another limitation is that although it is proposed that routine data could be enhanced using other datasets, only in-depth validation was undertaken using the

routine HID (the HES-equivalent). However, it is proposed that more events could be identified using additional sources.

Limitations were also identified regarding the reference data. Events can be missed from manually reviewing clinical noting data, for example if the event occurred in another hospital without transfer of the patient noting. A further limitation was that the radiotherapy reference data documented radiotherapy fractions that were prescribed and not those that were delivered. In contrast, the HID identified events that had been delivered and not prescribed. I queried the relationship between fractions prescribed and delivered to see if this would have implications on the accuracy results. Upon query, I found anecdotally at UHB QEH, the relationship between the fractions prescribed and delivered is extremely close, therefore implying that there would have been little impact on the sensitivity values identified. However, even if misclassified, the impact of this was deemed small. The sensitivity would increase if regimens that were thought to be missed in the routine data, due to being documented as being prescribed in the radiotherapy data, were never actually delivered; the number of false negatives in the routine data would reduce, increasing the sensitivity.

In summary, limitations were again identified for using routine data for RCTs, including missingness, inaccurate coding and limited clinical variables, and it is therefore of no surprise that previously there has been hesitation to using such a resource and therefore, limited evidence of such. However, I propose many of these limitations can be overcome.

5.7.3 Implications to RCT conduct

As presented in the results (5.6.4), *directly* using coding to identify events for a trial using the routine hospital administration data alone, was not feasible and as stated, this can be reflected by the lack of literature for trials being performed solely using routine healthcare data. However, as proposed (section 5.6.4) various techniques may enable the feasibility of this, 1) additional routine datasets, 2) a querying framework and 3) additional rules.

The aim was that additional datasets that have data derived from alternative sources would enhance the collection of events. A querying framework would enable the screening of a broader coding set, therefore, enabling the capture of additional but confirmed events. Additional rules may also enable the detection of missing routine data events that were still not identifiable using additional datasets or captured through weaker coding restrictions. These three techniques are detailed below. The BladderPath trial data collection framework is currently being designed to incorporate these techniques, with the aim to be regarded as the first secondary care or oncology RCT to use routine data as the basis for data collection.

5.7.3.1 Additional datasets

To aim to capture the maximum number of events, I proposed additional datasets would be required. In particular, radiotherapy events were of the highest concern due to the historically poor accuracy seen in figure 43. Hence, to further ensure that radiotherapy events could be captured nationally across multiple sites, I proposed the supplementation of the hospital interactions data (for example, HES) with data from the national radiotherapy dataset (RTDS). These data are collected directly from the radiotherapy treatment machines and these data are therefore derived from an alternative source to the EHR generated HES data (introduction 1.2). The RTDS is proposed to be automatically generated upon administration of radiotherapy (24).

In this study the RTDS-equivalent was used as a reference to validate the quality of the HES-equivalent. This was due to the hypothesised high quality because of the automatically generated method of collection. However, the RTDS-equivalent was also investigated by performing survival analyses with data from the Office of National Statistics (ONS) Spine (234) (death data). Note that A. Dosanjh performed this separate analysis, hence the results are not shown. This work undertook analyses of the impact of radiotherapy on overall survival in a bladder cancer cohort (utilising many of the patients within the cohorts presented during this study). The results generated by A.Dosanjh during the analyses revealed results comparable to published clinical trials results, validating the utility of these data for trial analyses.

In addition to enhancing HES-detected (27) radiotherapy events with the RTDS (235), the HES chemotherapy data can also be supplemented with the systemic anti-cancer therapy data set (SACT) (52) and imaging data can be supplemented with the Diagnostic Imaging Dataset (DID) (236). Cancer registration data (49) can also be requested to further enhance general event detection and also date of death (for overall survival calculation) can be requested (table 77). Additional datasets, derived from alternative sources, are proposed to enhance trial data quality and reduce missingness (table 77).

Outcome to be extracted	Database	Implication from validation
Surgery to bladder	<ul style="list-style-type: none"> Hospital Episode Statistics (HES) 	<ul style="list-style-type: none"> Historically high quality HES data alone sufficient
Chemotherapy regimens	<ul style="list-style-type: none"> Hospital Episode Statistics (HES) Systemic Anti-Cancer Therapy (SACT) 	<ul style="list-style-type: none"> Historically high quality to detect regimens The exact date of administrations can additionally be found in the SACT data (and clinical noting if required) HES data alone sufficient
Radiotherapy regimens	<ul style="list-style-type: none"> Hospital Episode Statistics (HES) National Radiotherapy Data set (RTDS) 	<ul style="list-style-type: none"> More recent high quality (since 2014) to detect regimens Due to the validation of the radiotherapy data, the RTDS could be used to supplement missing events
Cystoscopy	<ul style="list-style-type: none"> Hospital Episode Statistics (HES) Diagnostic Imaging Data set (DID) 	<ul style="list-style-type: none"> Recent high quality (since 2016) Consistent high quality TURBT coding Historically lower quality of flexible cystoscopy coding Prior to trial data confidence, a database query process may be necessary (check flag = if no flexible cystoscopy is identified prior to TURBT) To confirm identification of subsequent surveillance flexible cystoscopy events, the DID could be used as a supplement
BCG regimens	<ul style="list-style-type: none"> Hospital Episode Statistics (HES) Systemic Anti-Cancer Therapy (SACT) 	<ul style="list-style-type: none"> More recent high quality (since 2013) SACT data could supplement missing administration details
Censor	<ul style="list-style-type: none"> Hospital Episode Statistics (HES) 	<ul style="list-style-type: none"> Data quality historically high, but reduced recently (post-2016) Therefore, upon query at site, the most recent event in the clinical noting could be confirmed
Date of death (overall survival)	<ul style="list-style-type: none"> Various sources, including, the ONS spine and PHE NCRAS data 	<ul style="list-style-type: none"> The utility of such was demonstrated by A.Dosanji using data from the ONS Spine The collection of death data is mandatory and as such data quality should be high (to be confirmed)

Table 77: An example for the BladderPath trial of how additional datasets can aid event identification.

I have also submitted this table in a submitted paper pending publication.

However, a limitation to this approach is that the more data sets that are required, the more resources are needed to: 1) apply for these data, 2) safely receive these data (potentially from multiple routine data providers), 3) merge these data, 4) validate these

data and 5) process these data. More data sets therefore require additional time to undertake the steps listed above and hence, increase the costs associated with using routine data follow-up. However, the question remains as to if the costs associated with these methods remain lower than the costs associated with standard follow-up techniques.

Additional datasets will only help where the outcomes required by the trial are obtainable in these data. As presented in chapter 3 and 4, the routine data coding can be restrictive, including with the lack of clinical or pathological details. Outcomes are only feasible to be collected for trials if the particular outcomes of interest are *directly* identifiable (such as presented here) or *indirectly* identifiable (such as in chapter 4).

5.7.3.2 Querying framework

As mentioned above, in order to capture as many trial events as possible a methodology was developed which involved the querying of the algorithmically detected events (figure 44). Due to this proposed querying technique, the algorithm was developed to identify a broad coding list (as many events as possible), which could then be confirmed at trial sites. Hence, a very high sensitivity is required to detect as many events as possible, but a lower specificity would be acceptable. This is because each individual event would be queried at site. Hence all false positives would be removed, so by definition the specificity would become 100%. Therefore, a lower PPV would be acceptable, as the benefits from identifying additional events for an RCT outweighs the additional burden on site staff from validating further (potentially false positive) events.

A balanced approach will be required using this method; the broader the acceptable coding, the higher the burden on trial staff to query these additional events. Care also has to be taken with the documentation of the clinical noting. This is to ensure that routine data true positive events are not rejected (upon query) due to being missed in the clinical noting.

5.7.3.3 Additional rules

I proposed that additional rules could be utilised to flag missing events of interest. For example, I identified that flexible cystoscopies could be missed from the routine data

and therefore could be flagged using rules (cystoscopy, table 77); if a TURBT is identified, but no flexible cystoscopy prior to this, a flexible cystoscopy event could be queried in the clinical noting. This would enable identification of initial flexible cystoscopies, however, not later surveillance events. Similar rules could be adapted to increase the sensitivity of detecting the last censor events.

Limitations include the potential for misclassification, increasing site staff burden (103) from querying the additional events. In addition, further rule creation increases the complexity of the data collection and therefore may increase resource use and cost.

However, even in the absence of these additional methods to enhance the data integrity, the routine informatics data were seen to be of recent high quality for trial conduct; a level of missingness is also expected using conventional trial data (149, 237). Therefore, data quality is estimated to have minimal impact on the trial data integrity, especially if these additional methods are employed.

5.7.4 Further considerations for using routine data in trials

There are further considerations and potential caveats for using routine data in trials which must be considered whilst developing a routine data framework. Timeliness of accessing routine data is a limitation of using such methods in trials. This was illustrated during the data validation analyses. The three missing radiotherapy events in 2018 were not due to erroneous data but due to the delay in data access. Often data are provided at a delay and therefore the data censor limits identification of the most recent events. If the time lag between the event occurring and the routine data being received by the trial is too great, techniques such as these proposed here are futile.

However, I propose, although a time lag would be present through acquiring routine data, the time lag may be overall less than with standard collection methods. For example, in the STAMPEDE trial, initially follow-up is undertaken every six-weeks but at six years follow-up can occur at less frequent intervals of 12-months (72). Hence, using STAMPEDE as an example, obtaining data from routine providers, would reduce the mean delay in event detection over the whole follow-up period. Discussion is currently underway with the routine data providers to identify the optimum frequency and delay

in data release (by the data provider) for the lead BladderPath site to receive these data; the frequency is proposed to be 12-weekly and the delay in data collection is currently under discussion. Hence, in terms of data frequency, in comparison to the STAMPEDE trial, initially there would be a 6-week delay in event collection using the routine data but from 9-months, where follow-up become less frequent, the timeliness of event collection would increase using the routine data, compared to the standard methods. However, this is also dependent on the delay in data release by the data provider.

Consent and access to data is of further consideration when using these data in trials. This study was undertaken using non-trial audit patients; hence, consent was not required. However, for a trial, patients are required to consent to routine data access, which enables the acquisition of personally identifiable event data. The patient consent forms for the BladderPath trial were designed to enable this, including clauses such as (excerpts taken directly from the BladderPath consent forms):

- *I give permission to the Trials Office to access relevant data collected via the NHS Hospital Episode Statistics (HES). I understand this may include relevant data prior to the date of consent. I give permission for my personal details e.g. full name, date of birth, gender and NHS number to be sent to the University Hospitals Birmingham NHS Foundation Trust Informatics Team via a secure system to be able to obtain the relevant data. I understand that this transfer of my personal details will only occur once.*
- *I understand that the Trials Office, may access information held by Cancer Registries, Cancer Intelligence Unit, NHS Digital, National Cancer Registration and Analysis Service (NCRAS) and other similar data sources kept by the NHS or related organisations, to keep in touch with me and to follow-up on my health status.*

I have been leading the acquisition of the routine data for the trial. Initially an application was submitted to NHSD to acquire the HES, DID and ONS data and I was completing, in parallel, an application to PHE for the SACT, RTDS and cancer registration data. However, limitations of acquiring data from NHSD were present, including costings. This technique of using routine data for follow-up aims to reduce the cost of

follow-up compared to standard trial data collection methods. However, if the costs are too high to receive frequent data sets from these providers, then techniques such as these become futile. Where providers charge the same data linkage fee per extract, the costs can make these methods non-accessible by trials (see further details in 6.1.1.3).

Due to reasons such as costings, an application was being completed to hope to acquire all data from PHE. However, due to the importance of receiving timely data, discussions remain with NHSD and hence, a combination of acquiring data from both PHE and NHSD may be the optimum.

As discussed, due to the querying methodology proposed, all events will be queried for accuracy. Therefore, I am currently liaising with PHE regarding the development of a continual data quality feedback mechanism. The overall aim is to enhance routine data quality for future trials, where the requirement for querying events would be negated. Working closely with the routine data providers during trial set up is necessary to design a smooth mutually beneficial application.

It is vital that these caveats are taken into consideration whilst designing a routine data framework. However, if these concerns can be resolved, I believe in certain settings, routine data may be able to enhance the conduct of clinical trials.

5.8 Conclusion

Although clinical trials have used routine data to supplement standard data follow-up methods for many years, I believe there to be limited evidence of using routine data as the basis to conduct a UK trial. Once the methodology has been implemented, I believe this will be the first non-primary care UK RCT and the first oncology UK RCT, using routine data with no standard clinic follow-up.

Due to the data limitations identified across all the thesis studies, a methodology was developed to ensure maximum data integrity. This enables us to continually monitor data quality throughout the trial and if the quality is seen to be reliable, the necessity for querying may be removed (or at least for trials in the future). I believe to have demonstrated the feasibility of this approach despite the current routine data considerations. I hope this novel methodology may impact how future trials are conducted. To reiterate, there are many proposed benefits to using these data and hence, implications to patients. This includes, reducing burden, saving resources, reducing costs, increasing efficiency, all whilst collecting high quality rapidly updateable easily auditable data, enabling real-time data monitoring. These implications aim to be tested in practice on receipt of the routine data within an RCT setting.

6 CHAPTER SIX: Overall discussion and conclusion

6.1 Discussion

The main aim of this PhD was to answer the question: '*Can routinely collected data be used to inform randomised controlled trial outcomes in oncology*'? To answer this, the feasibility of using routine data for trial follow-up was assessed and where possible clinically useable instruments were developed, to detect trial outcomes. To identify outcomes, methods were developed to *directly* and *indirectly* utilise the coding, to undertake retrospective trial analyses and in addition, assess the feasibility of using routine data for prospective trial follow-up.

Within the literature, I identified no United Kingdom (UK) randomised controlled trials (RCT) in secondary care or oncology which had been designed to conduct follow-up using routine data alone. Hence, I investigated the feasibility of this. In parallel, I explored if existing follow-up could be replaced or supplemented within an existing trial. Examples for this included: assessing the feasibility of detecting trial non-survival endpoints and assessing the feasibility of answering a question not previously possible with the standard trial data. There was limited evidence in the literature of routine healthcare data being used to identify non-survival endpoints within trials (165, 171) and no evidence within the UK. In addition, the research question that was not previously possible to answer using standard trial data was chosen due to lack of evidence in the literature from a single RCT.

Although the results of this thesis have been discussed individually within chapter 3, 4 and 5, this chapter discusses the thesis as a whole, including the implications of the results. Of necessity, the results were presented in a linear fashion, however, all chapters informed each study and were largely undertaken in parallel. Hence, despite individual chapters identifying different outcomes using alternative methods, the overlapping principles are discussed below. Cumulatively the results, I believe, enabled me to answer the research question.

In the introduction, I proposed the potential of such methods but also the concerns (table 1) that needed addressing in order to make clinically useable instruments for trial

follow-up. Hence, these are discussed below, in the context of the proposed thesis aims (section 1.8.1).

6.1.1 Routine data concerns

Both practical and regulatory concerns for using routine data for trial follow-up were explored in table 1. In practice, the true nature of these concerns and how these were overcome during these studies are presented below. The concerns included; outcome availability, data accuracy, data linkage, cost, timeliness, bias, patient privacy, security and consent and length of data retention (table 1). Data outcome availability and accuracy were the two greatest limitations which I identified during my studies, both of which are significant barriers to using these data in practice (63).

6.1.1.1 Routine data outcome availability

Firstly, all three studies were limited by the availability of outcomes in the routine data. In chapter 3, a standard neutropenic sepsis event outcome, such as febrile neutropenia, seen in the cross-study comparisons presented in the literature review, was not available in the routine data (Hospital Episode Statistics, HES). Hence, a proxy for the standard trial definition of sepsis events had to be created, utilising the coding that was available (admission for sepsis or neutropenic event). Proxies to classify the events by hormone-sensitivity (hormone-sensitive prostate cancer (HSPC), castrate-resistant prostate cancer (CRPC)) also had to be developed, using relative timing of events to classify the sepsis and chemotherapy events (table 21). When compared to the clinical noting (section 3.12.3), all events were correctly classified by hormone-sensitivity. However, the use of proxies was expected to lead to misclassified events during the other analyses. This is further discussed below, with regards to identifying a bladder cancer cohort. The use of proxies have implications in practice if exact definitions of events are required to be collected; for example, trial protocol specified definitions of sepsis events are required for CRF collection.

Although a different definition for toxicity events was identified using the routine data proxy, this model did enable comparison, by chemotherapy timing, within a single RCT population. This was not previously possible using the case report form (CRF) collected STAMPEDE data, as the outcome was not explicitly collected in the trial CRFs to perform

the new analyses. These data, therefore, enabled hypotheses to be analysed that were not previously envisaged when the CRFs were being designed. However, to reiterate, the results were not conclusive, for example, the study did include some patients prescribed both HSPC and CRPC chemotherapy regimens (partially paired data) and hence may have had implications on the rates in the CRPC group.

In chapter 4, the routine data (HES) could also not be used *directly* to identify trial outcomes. Trial-based disease outcomes, such as progression endpoints were not identifiable in the routine data due to the inadequate availability of key outcome data fields (section 4.3). Specific protocol defined progression outcomes cannot be identified in the HES data and therefore ‘time to events’ such as the progression-free survival (PFS) could not be derived *directly*. Hence, a model was created to *indirectly* infer when trial events of interest had occurred. I developed this model based upon the clustering of routine data cancer-related events, from the HES procedure and diagnostic coding (figure 30). The aim was to allow patterns of cancer-related coding to contribute to the model *directly*, to flag absent outcomes.

The routine data model was not able to detect individual protocol defined non-survival endpoints (PFS, metastases-free survival (MFS), failure-free survival (FFS)), but a composite was in fact being identified (section 4.10.4). The model identified the first peak in clinically relevant activity and thus the time to event calculations using the model constituted the ‘activity-free survival’ or AFS. This was defined to be the time to the first most intensive healthcare interaction period for a patient. Although the model was not able to identify the protocol defined non-survival endpoints *directly*, I proposed the AFS as an alternative novel non-survival endpoint, subject to further validation.

Standard non-survival endpoints such as the FFS, MFS and PFS present clinically different timepoints; some endpoints are less clinically relevant than others (for example, the FFS) but to function as a surrogate endpoint for overall survival, criteria for surrogacy are required (chapter 4, 4.9.5). The algorithm was identifying the more clinically relevant protocol defined outcomes (MFS, PFS), compared to the FFS (a less clinically relevant endpoint) (section 4.10.4). Although a novel endpoint was being

identified, similar treatment effects (hazard ratios) were seen in line with the protocol defined non-survival endpoints (table 60) (table 63) (table 66).

A model such as the above, has implications in practice; routine data could be used to perform trial time to event analyses, and the algorithm derived outcomes could be used to validate trial outcomes. This may supplement the trial with additional data that was previously lost to follow-up (due to trial under-reporting). There are also non-trial implications; the model could be used to perform audits on 'real-world' data. The aim would be to test new hypotheses prior to trial set up. In addition, this model could be used to compare trial to non-trial patients to test 'real-world' intervention effects and confirm that the same impacts occur in the non-trial setting. In addition, the models could be used to assess whether a trial result has been truly implemented in practice.

I propose that, alongside further validation, this endpoint could be used in addition to (and not in replacement of) the standard trial endpoints. I also propose that this routine data-based outcome may provide a particularly *clinically relevant* endpoint (section 4.11). However, I believe further validation of the utility of the AFS as a surrogate endpoint is strictly mandatory, for example, within other treatment settings. Examples of other settings include, patients being treated with abiraterone hormone therapy and a prostate cancer cohort with a low risk of progression. Upon further validation, the AFS may pose a novel way to collect outcomes data to negate, reduce or supplement the requirement of standard data collection methods.

Further to this, HES also intentionally did not document some investigations, such as simple radiography (X-rays), as opposed to complex radiography (for example, magnetic resonance imaging (MRI)). Hence, these were not available to detect in these data. This is likely due to the function of the administrative data, which does not seek to capture interactions of relatively low price.

The lack of detailed clinical outcomes in the routine data also had implications within chapter 5. Identifying a cohort of patients was problematic. Here, within a non-trial bladder cancer cohort, a muscle invasive bladder cancer subset (MIBC) was required for analyses. However, this MIBC cohort was not *directly* identifiable in these data; hence a

proxy outcome was developed using procedure patterns to identify the cohort to enable the timing between events to be calculated. Using this proxy, a large number of potential MIBC patients were missed (section 5.6.2). However, I believe this to have minimal impact within a standard RCT setting, if cohorts are being identified using standard methods. However, this is reflective of the potential for misclassified events using *indirect* proxy techniques.

In contrast, during the chapter 5 bladder cancer data quality validation study, the events of interest were *directly* available in these data. This included treatments and the last known follow-up indicator; hence, I developed a simple algorithm to extract these.

To summarise, to identify outcomes of interest in the routine data that were not present *directly* in these data, proxies had to be created to *indirectly* identify the outcomes. Novel event definitions were created, where possible; however, when novel outcomes could not be generated, when particular defined outcomes were required, it was clear that outcomes were being misclassified (for example, the MIBC cohort). In both chapter 3 and 4, novel outcomes were identified, to fit the outcome to these data to enable detection. Of particular interest is the model developed in chapter 4, based upon the clustering of events. These techniques appeared to enable trial event detection from routine data sources. However, whether these models work in practice, within a different population or setting are yet to be investigated and all events require validation on a study-specific basis. If the outcomes of interest are not available in the data sets, nor is it appropriate to utilise a proxy, then routine data would not be a suitable source of outcomes.

Summary implications for practice: Routine data outcome availability meant that some outcomes of interest were not directly available in these data. However, in this study proxies were developed to enable identification of such events. There is limited literature available that seek to extract clinically useable outcomes from routine data. For example, there is a lack of studies developing models to identify non-survival endpoints for RCT use (165, 171). I believe this to be due to the challenges highlighted during this research.

6.1.1.2 Routine data accuracy

Accuracy was investigated in depth throughout the studies. For example, during chapter 3, chemotherapy regimens were often not being coded for in the routine data and hence, not being detected. However, models were successfully created to account for the missing routine data chemotherapy regimens (inferred). This enabled detection of more true positive chemotherapy and sepsis events (for example, table 33). However, there was an increase in false positives that were unrelated to chemotherapy (table 27).

Further to this, to identify additional events of interest that were potentially miscoded, in all studies (except cystoscopy events within the chapter 5 time to correct treatment analyses), the eligible coding list to screen was broadened. This was to identify events that may have been documented using a similar, but incorrect, procedure code. Although enabling the identification of additional events of interest, this led to an increase in false positives (misclassification) being identified (for example, chapter 3 and 5); in chapter 3, admission events coded exclusively for neutropenia-only, may not have been infection admission events. In contrast, during chapter 5, the algorithm only found a small proportion of eligible MIBC patients for analyses. This is hypothesised to be due to the restrictive cystoscopy coding proxy. However, in the sample of the MIBC patients that were confirmed, 100% sensitivity for detection of events of interest was achieved. Despite the restrictions, false positive MIBC patients were also still identified but by confirming each event against note review these were excluded from the time to correct treatment analyses.

Further to this, it was identified that some HES procedures were coded with alternative descriptions to what was expected, for example, imaging events were coded as allied health professional consultations. Hence, although it would appear that events were missing, sometimes they were present but coded by alternative descriptions.

To identify further events, the routine administrative HES data were supplemented with additional datasets. This included the STAMPEDE data to identify HSPC chemotherapy regimens and the Systemic Anti-Cancer Therapy Dataset (SACT) data to identify chemotherapy events, including drug name. Linking the routine data to both the STAMPEDE and the SACT enhanced the identification of events. HES identified

admissions for infection events (sepsis plus or minus neutropenia coded events) with high accuracy when compared to the STAMPEDE data (figure 18, B).

In all three studies the false positive events identified using the routine data may have actually been true positives, that were unreported in the reference data sets (the clinical noting and the STAMPEDE trial data). Hospital clinical noting data may miss events occurring at other hospitals or events occurring prior to electronic recording of patient notes. Trial data are also known to suffer from loss to follow-up (under-reporting) and therefore some events could remain unreported, for example, due to recall bias, or patient relocation. Hence, routine data could be used to identify events missed using the standard trial data collection framework.

In chapter 4, HES was seen to miss prostate-cancer related events including skeletal-related events (SRE) (for example, table 57), due to the *direct* coding requirement. This is of utmost concern if the raw data are being used to collect such events. However, HES was able to identify outcomes that were missed in the trial data (for example, progression events, figure 31 and SREs, figure 32). This included identifying outcomes not required to be documented, but also some missed (section 4.10.3). Missing data in trials are a common feature but leads to great concern (149, 237). The HES detected outcomes were therefore able to complement the trial detected outcomes to enhance overall event collection. Although prostate-cancer related events were missed, the accuracy did increase over time (for example, identifying SREs in table 57). This increase in coding quality was reflected in the chapter 5 study, where the majority of the more recent events of interest were captured (table 75). In addition, the use of a clustering model to identify events in chapter 4, meant that single missing, wrongly coded or false positives would have minimal overall impact on the outcomes detected.

To reduce the impact of using broader event inclusion, where specific events were required to be detected, it was identified that it was necessary to query events (such as in the time to correct treatment analyses). Hence, to build a model to enable the feasibility of using routine data as the basis of RCT follow-up, a querying framework was designed (chapter 5). The principle was that all events of interest could be extracted, using a broader coding list, to capture further events but the false positives could then

be further excluded from the analyses. Within the RCT framework, each individual extracted routine data event, marked with a timestamp was proposed to be queried at the site where the event occurred, using the clinical noting (figure 44). Implications of this in practice include, both the identification of more follow-up events and also improved accuracy. There is potential that the clinical noting will be missing events, for example those occurring at other sites, and as such when the query is run from the routine data, the event could be rejected as a suspected false positive. If the event occurred at a different site, this is estimated to have little impact, as the query can be sent to the correct site based upon location codes in the routine data. However, incorrect clinical noting may lead to wrongly rejected events.

One aim of conducting a trial using routine data is to reduce burden on trial staff (103). However, event queries to the sites are necessary due to the data quality issues highlighted here. It is therefore vital to come to a compromise for maximum data quality but minimal staff burden. The ability to flag a suspected event to search for within the clinical noting is hypothesised to reduce the time taken to read all clinical noting entries, especially upon long-term follow-up. I have been trying to develop the most durable method for data flow, from receipt from the routine data provider to the validated data point being entered into the trial database. For example, consideration of temporal changes to the data are required. The data providers can continually send data updates for events already queried. Due to this, I am liaising with the data programming team to incorporate a timestamp into the database and a method to identify where data updates have been made to account for this. However, if the event has already been confirmed in the clinical noting data then any updates to the routine data may not be necessary to query again and duplicated events could greatly increase the burden on staff. In addition, the event coding may change when new coding classifications are released. Therefore, the algorithms and the trial database would have to be updated to enable the query of these new events (see discussion, section 4.11 for further details).

Due to the improvement in accuracy achieved, along with the additional processes developed for further integrity, I believe these data to be of sufficient accuracy for trial follow-up. As discussed, I am in the process of applying the framework developed in chapter 5 to the new trial, BladderPath, where I am part of the Trial Management

Committee. I am liaising with routine data providers to set up a continual data quality feedback mechanism. Due to querying every event of interest in the clinical noting, prior to being incorporated into the trial database, a direct measure of real-time data quality can be fed back to the routine data providers for service improvement. The aim is to enhance these data for future trials, where potentially the need for query will be negated.

Summary implications to practice: *Data accuracy was seen to be improving over time. However, techniques had to be developed to enhance the quality of the routine data to enable feasibility of use for trial data collection. Additional data processing techniques (such as those presented) are required before there is evidence of high-quality reliable coding. There is limited literature documenting the use of routine data for oncology RCT follow-up. This may be due to the accuracy concerns detailed within these studies.*

6.1.1.3 Other

In addition, there are other concerns for using routine data for trial conduct that were investigated during this project, for example, cohort linkage. I successfully completed a data application to Public Health England (PHE), in which STAMPEDE linkers were sent to the data provider so these data could be returned to us attached to the trial number. However, due to inaccurately documented trial linkers in the trial database, such as NHS number, a small proportion of patients failed to be linked. Errors can occur in the documentation of identification numbers, for example, NHS numbers. Upon telephone randomisation from the clinical trials unit to the site, these numbers can be misinterpreted. Hence, these patients were lost to follow-up using the routine data. I have informed the Medical Research Council Clinical Trials Unit at University College London (MRC CTU at UCL) that they need to query the identified inaccurate NHS numbers at site and prospectively collect these with more care. This is of utmost importance if follow-up is to be undertaken using routine data. In addition, this has highlighted how important it is to document linkers correctly for a trial using routine data as the main source of follow-up data; for example, the BladderPath trial. If the linkers are not documented correctly for BladderPath, these patients could be permanently lost to follow-up.

Another concern is the cost of routine data; costing is variable across providers. For example, NHSD charge a new linkage rate for each extraction, per dataset, of £2,060 (238) (in addition to other costs). If long-term follow-up at frequent intervals is required, as necessary in clinical trials, the use of these data are restricted. For example, a cost of £247,200 would be quoted for monthly follow-up over ten years, using only one linked dataset, just to acquire these data. This is in comparison to PHE sourced data, from anecdotal evidence, alongside the evidence gained from the STAMPEDE PHE application; it is suggested that PHE charge an initial linkage fee to develop an algorithm to extract and link these data. However, it is suggested that this fee is not re-billed for subsequent extractions. This can lead to a large difference in costing frameworks and can be the difference between feasibility and non-feasibility of this framework in a trial. I am liaising with the providers currently, for the BladderPath trial, to create an arrangement that may enable the trial to save resources and benefit from reduced running costs.

Regulatory concerns are also common. During this PhD I have successfully completed one data application and I have another underway. The completed application (STAMPEDE) was requesting retrospective data for patients that were randomised from when the trial began in 2005. This is in contrast to the pending application (BladderPath) which is prospectively requesting frequent drops of data. Both applications have crossed the timeline of GDPR initiation and therefore both applications had requirement changes during the application process. The benefit of the BladderPath design is that we are able to apply for these data in parallel to setting up the trial. This is in contrast to applying for retrospective data within the STAMPEDE trial. It is vitally important to liaise with routine data providers during the application (ideally during trial set-up) to enable the acquisition of data to be as quick and smooth as possible. It takes a long time to apply for routine data and therefore, as soon as the trial team decide they may like to use this data source, the application process should start.

Despite having successfully completed an application for routine data, when applying for retrospective trial data it is vitally important that the trial teams actively track those that do not give consent and withdraw from the trial; problems were experienced during the completion of this thesis. Patients that were identified initially as a consenting cohort, which were sent to the provider for linkage, were then actually identified as not having

given consent. These patients were excluded from future analyses and where possible removed from existing analyses, which were re-run. Once the MRC CTU at UCL reported these concerns, I frequently kept the team informed. I also informed the Warwick Medical School team who were happy with the measures taken. However, review of the patient consent is continually underway and additional patients may be identified that never consented. The MRC CTU at UCL are also currently seeking advice into what specific consent is required for such analyses. It is possible that due to the use of these data these particular studies may only require standard STAMPEDE trial follow-up consent, meaning only those withdrawing from the trial would be ineligible for the analyses. It is vitally important that trials remain vigilant with such consent data and efforts are being made by the trial team to enhance their documentation processes to increase integrity. Review of patient consent and enhancing the documentation of this is currently underway by the MRC CTU at UCL.

Routine data retention is also a concern; trials require data to be kept for audit purposes, often after trial closure. I am currently liaising with the routine data provider (PHE) for the BladderPath trial, to agree a length of time that the trial team can keep these data, however additional applications for retention can be made. We are currently in discussions with PHE to keep the data for ten years, however, during this time it will be necessary to apply for data retention.

Summary implications to practice: Other concerns such as linkage, cost, regulations, and data retention are limitations to utilising routine data in trials and are widely acknowledged (63). These concerns were encountered during the conduct of these studies. However, all were resolved in principle, and I therefore believe these concerns, previously believed to be barriers, can be overcome.

These limitations reflect why minimal research has been done into the feasibility of using routine data as the main source for trial follow-up. However, as described, frameworks can be developed to overcome these issues, to enable feasibility. Although I believe these studies to have shown great scope in using routine data, there is a potential that in practice, such resources are not yet ready for current clinical trial use. For example, although data quality has been seen to improve, there is a chance that the

quality could remain variable across different trusts and hence, not reduce burden on site staff. It is also possible that this technique may not reduce costs and that these data cannot be retrieved as soon as required. However, until trials are designed to test the feasibility of this, we will not know.

6.1.2 Routine data trial potential

There are also many potentials for using routine data to conduct trials (introduction section 1.7, table 3). These are discussed in relation to the results presented in this thesis. These include the potential of investigating new hypotheses, event validation, identifying further events for example, those lost to follow-up (63), reducing burden (103), reducing costs (63), and increasing timeliness (63), which all have potential impacts on patients, discussed below.

6.1.2.1 New hypotheses

New hypotheses not previously possible to answer using standard trial data were investigated. In chapter 3 the routine data mitigated the need for cross-study comparisons and enabled the new clinical rationale to be investigated using a single randomised group. In addition, the routine data outcomes identified in chapter 4 were developed with the purpose of answering clinical hypotheses not previously possible with standard trial data. These include those such as, investigating time to subsequent events. Only the first event is required to be collected by STAMPEDE, so this would enable comparisons of subsequent treatment responses. In addition, in chapter 5, the routine data enabled us to use a non-randomised population to identify MIBC treatment outcomes.

However, due to the limitations of the routine data described above, routine data would not be appropriate to answer all new hypotheses. The data restrictions mean that both suitable and feasible research questions have to be proposed. For example, I am currently working on a project that seeks to investigate cardiovascular events on hormone therapy. These events are *directly* available in these data and thus I assessed that routine data analyses would seem a feasible method to undertake this.

6.1.2.2 Validation of trial events and identifying events lost to follow-up

The routine data were also able to validate trial collected events. HES validated STAMPEDE neutropenic events (chapter 3) and trial endpoint events (chapter 4). In addition, more events could be identified using routine data than were identified using the standard trial data. In chapter 3, additional neutropenic events were identified than in the trial, confirmed by note review. However, these were events that were not required to be collected by the trial. However, in the large analyses a low PPV of HSPC events was identified, compared to the HSPC STAMPEDE events. It is proposed that many of these events may actually be true positives, missed by the trial. Therefore, I am in discussion with the MRC CTU at UCL to confirm a sample of these events for verification at individual sites. Further to this, in chapter 4, outcomes were identified that were missed by the trial, as confirmed in the note review. Some were missed due to loss to follow-up and others due to the requirement for STAMPEDE to only document the first outcome. I recommend that any additional routine data identified events, that are potentially missed by the trial, should be queried at the individual sites and upon confirmation, subsequently be added to update the trial database. It is also possible that the proxy model developed in chapter 4 could be used to flag at risk patients, for example, those entering a palliative care stage. Therefore, I believe, routine data can be used to validate or enhance trial data.

In terms of maximising clinical trial data event collection, I believe that the use of routine data detected outcomes to supplement trial detected outcomes is a superior method of data collection. Aside from the potential for increased resource use by using both methods of collection, I believe that the additional events found by the routine data models could be used to enhance outcome collection (once these events have been validated at site).

6.1.2.3 Other

The benefits of reduced costs, burden and timeliness of collecting events (63) are yet to be confirmed upon the acquisition of the BladderPath routine data. However, the proposed framework is hypothesised to aid these, when compared to standard patient-health care professional follow-up.

I have presented here studies investigating the feasibility of using routine data to identify oncology trial outcomes. I investigated the ability to reduce/replace standard follow-up and the use of routine data as the basis of follow-up. One overarching limitation to this study was that the algorithms (chapter 3, 4 and 5) were developed using small cohort single-site data only. The model in chapter 4 (proxy endpoint) was validated, where possible, using multi-site trial data but for chapter 3 (neutropenic events study) and 5 (the BladderPath framework) multi-site validation was not possible. Where clinical results were presented (chapter 3), these should be interpreted with caution, are inconclusive and require further validation. As discussed previously, the algorithms may respond differently to data generated from different sites, impacting the ability to detect events and therefore the integrity of the algorithms.

If I were to choose one result from this study to present to funders to persuade them to use in practice, I would choose figure 38. Although, the novel surrogate endpoint developed was only investigating proof-of-concept, I believe that there is potential that we may be able to isolate a routine data-based proxy that may correlate with standard clinical trial outcomes. I also believe that this may have the biggest impact on patient care. The clinical trial framework may be enhanced if events missed using standard methods could be identified, if at risk patients could be flagged, if an alternative method to collect time-to-event data was available (that may correlate with quality of life) and if methods could enable analyses of real-world treatment effects. I believe all of these could impact patients directly, hence, I chose this figure.

I believe this study to have shown that routine data are a feasible data source for trial follow-up; however, extensive quality control techniques are required. I also believe standard trial data collection methods and routinely collected healthcare data to be complementary and their symbiotic relationship enables an enhanced trial framework.

I believe that a routine data framework can aid the analyses of new hypotheses, enable trial outcome validation and identify outcomes lost to follow-up. Potential implications may include: reduced burden on health care professionals and patients (103, 104), reduced costs (63, 103), resources (63, 103) and increased timeliness of collecting trial events (63). Further potential uses of these data include cross comparing treatments to

real world outcomes to identify how treatments behave in a non-trial population. The final implication of utilising such techniques would be, enhanced trials, enabling more interventions to be available to patients, to enhance length and quality of life.

6.2 Future work

Future work is required that was beyond the scope of this thesis. The 'activity-free survival' (AFS) outcome rules need to be enhanced for use in an alternative setting to that in which it was designed. The optimum thresholds and weekly intervals need to be re-confirmed. The model may work differently with different classes of treatment; hence, further validation is required. I have also discussed with various organisations, such as, pharmaceutical companies, universities and hospitals, the validation of the AFS outcome in different oncology settings. Also, within the same prostate cancer setting, I have discussed testing the model in a low risk (of progression) prostate cancer trial. I am currently in discussions with a pharmaceutical company regarding potential projects, such as, assessing how the deoxyribonucleic acid (DNA) damage repair status may affect the response to chemotherapy, using routine models such as those outlined in this thesis.

The model is also planned to be used to answer hypotheses not possible using the protocol collected STAMPEDE data, such as response to subsequent treatments. A PhD proposal is currently being developed which includes enhancing the algorithm developed here and using it to answer other clinical questions. I am having input into the research proposal and suggesting various techniques that were sadly beyond the timescale of this thesis. These include, additional modelling techniques.

New protocols are currently being written to answer new hypotheses using the routine data I acquired. For example, assessing fractures and cardiovascular events whilst on abiraterone hormone therapy within the STAMPEDE trial; I am involved with this work.

In addition, the data collection framework, presented in chapter 5, is currently being implemented in the BladderPath trial. Hence, the feasibility in practice is yet to be determined, including the overall impact on efficiency and costs. The applicability in other trials in and outside of oncology also needs to be tested.

6.3 Conclusion

Routinely collected healthcare data are a feasible and potentially useful data source for trial conduct. During this proof of principle study, a method was developed to identify a novel surrogate endpoint for trial effectiveness studies. I believe this to be the first oncology routine data non-survival endpoint developed, and the first instance where these outcomes have been used to perform trial treatment effectiveness analyses. In addition, I believe this thesis to present a novel framework developed to perform oncology trial analyses using routine data as the basis of follow-up. I believe this to have not previously been done in an oncology or secondary care setting in the United Kingdom (UK).

However, the integrity and clinical feasibility of these models need further validation on a trial by trial basis, as many limitations were identified during this proof of principle study. Despite this, methods were developed with the aim to overcome these limitations, to produce clinically useable tools. I believe routine data have potential to increase trial efficiency, with the aim to have access to more patient interventions across the whole of healthcare. As routine data continues to improve, perhaps in the future routine data led trial follow-up may become the standard.

Finally, *can routinely collected data be used to inform randomised controlled trial outcomes in oncology?*

Yes. But...be mindful

7 **References**

1. Mintz HP, Evison F, Parsons HM, Sydes MR, Spears MR, Patel P, et al. National, centralised hospital datasets can inform clinical trial outcomes in prostate cancer: A pilot study in the STAMPEDE trial. *Journal of Clinical Oncology*. 2017;35(6 Supplement 257).
2. Academy of Medical Sciences and the Association of the British Pharmaceutical Industry. Looking to the future: oncology endpoints (2017) - <https://acmedsci.ac.uk/file-download/41135280> [Accessed: 25 Jan 2018].
3. Arruebo M, Vilaboa N, Sáez-Gutierrez B, Lambea J, Tres A, Valladares M, et al. Assessment of the evolution of cancer treatment therapies. *Cancers*. 2011;3(3):3279-330.
4. Thornton J. Some cancers are “missing out” on new drugs. *British Medical Journal*. 2018;363:k5128.
5. Collier R. Rapidly rising clinical trial costs worry researchers. *Canadian Medical Association Journal*. 2009;180(3):277-278.
6. Thomas D. *Clinical Pharmacy Education, Practice and Research: Clinical Pharmacy, Drug Information, Pharmacovigilance, Pharmacoeconomics and Clinical Research*: Elsevier; 2018.
7. Griggs RC, Batshaw M, Dunkle M, Gopal-Srivastava R, Kaye E, Krischer J, et al. Clinical research for rare disease: opportunities, challenges, and solutions. *Molecular genetics and metabolism*. 2009;96(1):20-6.
8. Hay M, Thomas DW, Craighead JL, Economides C, Rosenthal J. Clinical development success rates for investigational drugs. *Nature biotechnology*. 2014;32(1):40.
9. Sydes MR, Parmar MKB, James ND, Clarke NW, Dearnaley DP, Mason MD, et al. Issues in applying multi-arm multi-stage methodology to a clinical trial in prostate cancer: the MRC STAMPEDE trial. *Trials*. 2009;10(1):39.
10. James S, Rao SV, Granger CB. Registry-based randomized clinical trials-a new clinical trial paradigm. *Nature Reviews Cardiology*. 2015;12(5):312.
11. Lewsey JD, Leyland AH, Murray GD, Boddy FA. Using routine data to complement and enhance the results of randomised controlled trials. *Health Technol Assess*. 2000;4(22):1-55.

12. Gantz J, Reinsel D. Extracting value from chaos. IDC iview. 2011;1142(2011):1-12.
13. Dinov ID. Volume and value of big healthcare data. Journal of medical statistics and informatics. 2016;4:3.
14. Zhang Z. Big data and clinical research: perspective from a clinician. Journal of thoracic disease. 2014;6(12):1659-64.
15. Bellazzi R, Zupan B. Predictive data mining in clinical medicine: current issues and guidelines. International journal of medical informatics. 2008;77(2):81-97.
16. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health Information Science and Systems. 2014;2(1):3.
17. One institute strategy 2019/20. Health data research UK (HDR UK) [ONLINE] Available at: <https://www.hdruk.ac.uk/wp-content/uploads/2019/07/190606-HDR-UK-One-Institute-Strategy-compressed-for-website.pdf> [Accessed: 05 Aug 2019].
18. Better, faster and more efficient clinical trials. Health data research UK (HDR UK) [ONLINE] Available at: <https://www.hdruk.ac.uk/research/research-priorities/21st-century-clinical-trials/> [Accessed: 05 Aug 2019].
19. Data Resources in the Health Sciences. Health Sciences library [ONLINE] Available at: <http://guides.lib.uw.edu/hsl/data/findclin> [Accessed: 08 July 2019].
20. Murray MD. Use of data from electronic health records for pharmacoepidemiology. Current Epidemiology Reports. 2014;1(4):186-93.
21. Scott PJ, Curley PJ, Williams PB, Linehan IP, Shaha SH. Measuring the operational impact of digitized hospital records: a mixed methods study. BMC medical informatics and decision making. 2016;16(1):143.
22. Ford E, Carroll JA, Smith HE, Scott D, Cassell JA. Extracting information from the text of electronic medical records to improve case detection: a systematic review. Journal of the American Medical Informatics Association. 2016;23(5):1007-15.
23. Blumenthal S. Improving Interoperability between Registries and EHRs. AMIA Summits on Translational Science Proceedings. 2018;2017:20.
24. Hoskin P, Forbes H, Ball C, Riley D, Cooper T. Variations in radiotherapy delivery in England—evidence from the National Radiotherapy Dataset. Clinical Oncology. 2013;25(9):531-7.

25. The clinical practice research datalink (CPRD) [ONLINE] Available at: <https://www.cprd.com> [Accessed: 08 July 2019].
26. Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, et al. Data resource profile: clinical practice research datalink (CPRD). *International journal of epidemiology*. 2015;44(3):827-36.
27. Hospital Episode Statistics. NHS Digital [ONLINE] Available at: <http://content.digital.nhs.uk/hes> [Accessed: 12 Feb 2018].
28. The SAIL databank [ONLINE] Available at: <https://saildatabank.com> [Accessed: 08 July 2019].
29. Ford DV, Jones KH, Verplancke J-P, Lyons RA, John G, Brown G, et al. The SAIL Databank: building a national architecture for e-health research and evaluation. *BMC health services research*. 2009;9(1):157.
30. Information services division (ISD). ISD Scotland [ONLINE] Available at: <https://www.isdscotland.org> [Accessed: 08 July 2019].
31. Evans JM, MacDonald TM. Record-linkage for pharmacovigilance in Scotland. *British journal of clinical pharmacology*. 1999;47(1):105-10.
32. ICD Classifications. World Health Organisation (WHO) [ONLINE] Available at: <http://www.who.int/classifications/icd/en/> [Accessed: 22 June 2018].
33. OPCS Classification of Interventions and Procedures. National Health Service (NHS) [ONLINE] Available at: https://www.datadictionary.nhs.uk/web_site_content/supporting_information/clinical_coding/opcs_classification_of_interventions_and_procedures.asp?shownav=1 [Accessed: 04 Sept 2019].
34. Pazdur R. Endpoints for assessing drug activity in clinical trials. *The oncologist*. 2008;13(Supplement 2):19-21.
35. Fitzpatrick T, Perrier L, Shakik S, Cairncross Z, Tricco AC, Lix L, et al. Assessment of Long-term Follow-up of Randomized Trial Participants by Linkage to Routinely Collected Data: A Scoping Review and Analysis. *Journal of the American Medical Association (JAMA) network open*. 2018;1(8):e186019-e.
36. Herbert A, Wijlaars L, Zylbersztejn A, Cromwell D, Hardelid P. Data resource profile: hospital episode statistics admitted patient care (HES APC). *International journal of epidemiology*. 2017;46(4):1093-i.

37. The quality of clinical coding in the NHS [ONLINE] Available at: http://www.chks.co.uk/userfiles/files/The_quality_of_clinical_coding_in_the_NHS.pdf [Accessed: 04 Sept 2019].
38. Farrar S, Yi D, Sutton M, Chalkley M, Sussex J, Scott A. Has payment by results affected the way that English hospitals provide care? Difference-in-differences analysis. *British Medical Journal*. 2009;339.
39. Users, uses and access to Hospital Episode Statistics. NHS Digital. [ONLINE] Available at: <https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics/users-uses-and-access-to-hospital-episode-statistics> [Accessed: 08 July 2019].
40. Hughes GJ, Van Hoek AJ, Sriskandan S, Lamagni TL. The cost of hospital care for management of invasive group A streptococcal infections in England. *Epidemiology & Infection*. 2015;143(8):1719-30.
41. Luengo-Fernandez R, Silver LE, Gutnikov SA, Gray AM, Rothwell PM. Hospitalization Resource Use and Costs Before and After TIA and Stroke: Results from a Population-Based Cohort Study (OXVASC). *Value in Health*. 2013;16(2):280-7.
42. Haslam N, Hoang U, Goldacre MJ. Trends in hospital admission rates for whooping cough in England across five decades: database studies. *Journal of the Royal Society of Medicine*. 2014;107(4):157-62.
43. Keenan TDL, Yeates D, Goldacre MJ. Uveal melanoma in England: trends over time and geographical variation. *British Journal of Ophthalmology*. 2012;96(11):1415-9.
44. Hippisley-Cox J, Coupland C. Predicting risk of upper gastrointestinal bleed and intracranial bleed with anticoagulants: cohort study to derive and validate the QBleed scores. *British Medical Journal*. 2014;349:g4606.
45. Billings J, Dixon J, Mijanovich T, Wennberg D. Case finding for patients at risk of readmission to hospital: development of algorithm to identify high risk patients. *British Medical Journal*. 2006;333(7563):327.
46. Thorn JC, Turner E, Hounscome L, Walsh E, Down L, Donovan J, et al. Validation of The Hospital Episode Statistics Outpatient Dataset in England. *Value in Health*. 2014;7(17):A547-A8.
47. Liu ZW, Fitzke H, Williams M. Using routine data to estimate survival and recurrence in head and neck cancer: our preliminary experience in twenty patients. *Clinical Otolaryngology*. 2013;38(4):334-9.

48. The HES processing cycle and data quality v4.0. NHS Digital [ONLINE] Available at: <https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics/the-processing-cycle-and-hes-data-quality> [Accessed: 04 June 2019].
49. National cancer registration and analysis service (NCRAS). Public Health England [ONLINE] Available at: <http://www.ncin.org.uk/home> [Accessed: 28 June 2019].
50. Cancer outcomes and services dataset (COSD) [ONLINE] Available at: http://www.ncin.org.uk/collecting_and_using_data/data_collection/cosd [Accessed: 28 June 2019].
51. Pickering L, Larkin J. Systemic anti-cancer therapy (SACT) dataset. The Lancet Oncology. 2014;15(10):1063.
52. Systemic Anti-Cancer Therapy dataset. National Cancer Registration and Analysis Service [ONLINE]. Available at: http://www.ncin.org.uk/collecting_and_using_data/data_collection/chemotherapy [Accessed: 23 Aug 2018].
53. Bright CJ, Lawton S, Benson S, Bomb M, Dodwell D, Henson KE, et al. Data Resource Profile: The Systemic Anti-Cancer Therapy (SACT) Dataset. International journal of epidemiology. 2019.
54. Hunt K et al. National Cancer Registrars Association. Cancer Registry Management: Principles and Practice. 2004.
55. Wiederhold G. Future of security and privacy in medical information. Studies in health technology and informatics. 2002;80:213-29.
56. Steward M. Electronic medical records: Privacy, confidentiality, liability. The Journal of legal medicine. 2005;26(4):491-506.
57. Barrows RC, Clayton PD. Privacy, confidentiality, and electronic medical records. Journal of the American Medical Informatics Association. 1996;3(2):139-48.
58. Caldicott C. The Caldicott Report. Department of Health. 1997.
59. McCall B. What does the GDPR mean for the medical community? The Lancet. 2018;391(10127):1249-50.
60. The Information Commissioner's Office (ICO). The principles [ONLINE] Available at: <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/> [Accessed: 04 Sept 2019].

61. John Wilbanks: Let's pool our medical data. [ONLINE] Available at: https://ted2srt.org/talks/john_wilbanks_let_s_pool_our_medical_data [Accessed: 28 June 2019].
62. The Times online. Don't let privacy concerns halt a health revolution. [ONLINE] Available at: <https://www.thetimes.co.uk/article/don-t-let-privacy-fears-halt-a-health-revolution-z2jkcknh8> [Accessed: 04 Sept 2019].
63. Mc Cord KA, Salman RA-S, Treweek S, Gardner H, Strech D, Whiteley W, et al. Routinely collected data for randomized trials: promises, barriers, and implications. *Trials*. 2018;19(1):29.
64. Lee F, Patel HRH, Emberton N. The 'top 10' urological procedures: a study of hospital episodes statistics 1998-99. *Bju International*. 2002;90(1):1-6.
65. Ballaro A, Oliver S, Emberton M. Do we do what they say we do? Coding errors in urology. *BJU international*. 2000;85(4):389-91.
66. Dettori J. Loss to follow-up. *Evidence-Based Spine-Care Journal*. 2011;2(1):7-10.
67. 19/49 call for ambitious data-enabled trials, health services and public health research studies. NIHR [ONLINE] Available at: <https://www.nihr.ac.uk/funding-and-support/documents/current-funding-opportunities/data%20enabled%20trials/19-49-Data-Enabled-Trials-finalv2.pdf> [Accessed: 08 July 2019].
68. Weir HK, Anderson RN, King SMC, Soman A, Thompson TD, Hong Y, et al. Peer Reviewed: Heart Disease and Cancer Deaths—Trends and Projections in the United States, 1969–2020. *Preventing Chronic Disease*. 2016;13:160211.
69. Townsend N, Wilson L, Bhatnagar P, Wickramasinghe K, Rayner M, Nichols M. Cardiovascular disease in Europe: epidemiological update 2016. *European heart journal*. 2016;37(42):3232-45.
70. Center MM, Jemal A, Lortet-Tieulent J, Ward E, Ferlay J, Brawley O, Bray F. International variation in prostate cancer incidence and mortality rates. *European Urology*. 2012;61(6):1079-92.
71. Antoni S, Ferlay J, Soerjomataram I, Znaor A, Jemal A, Bray F. Bladder cancer incidence and mortality: a global overview and recent trends. *European Urology*. 2017;71(1):96-108.
72. Systemic therapy in Advancing or Metastatic Prostate Cancer: Evaluation of Drug Efficacy, A multi-arm multi-stage randomised controlled trial. The protocol v19 [ONLINE] Available at:

http://www.stampetrial.org/media/2032/stampede_protocol_v190_clean.pdf
[Accessed: 05 Sept 2019].

73. The BladderPath trial protocol v3 [ONLINE] Available at: <https://www.birmingham.ac.uk/Documents/college-mds/trials/crcu/BladderPath/Current/protocol/BladderPath-Protocol-v3.0-19-Feb-2019.pdf> [Accessed: 05 July 2019].

74. Prostate cancer statistics. Cancer Research UK [ONLINE] Available at: <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/prostate-cancer#heading-Zero> [Accessed: 22 Aug 19].

75. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *International journal of cancer*. 2015;136(5):E359-E86.

76. Advanced prostate cancer. Prostate cancer UK. [ONLINE] Available at: <https://prostatecanceruk.org/prostate-information/just-diagnosed/advanced-prostate-cancer> [Accessed: 18 July 2019].

77. Créhange G, Chen CP, Hsu CC, Kased N, Coakley FV, Kurhanewicz J, et al. Management of prostate cancer patients with lymph node involvement: a rapidly evolving paradigm. *Cancer treatment reviews*. 2012;38(8):956-67.

78. Drake CG. Visceral metastases and prostate cancer treatment: 'die hard,' 'tough neighborhoods,' or 'evil humors'? *Oncology*. 2014;28(11):974.

79. If your prostate cancer comes back. Prostate cancer UK [ONLINE] Available at: <https://prostatecanceruk.org/prostate-information/treatments/if-your-prostate-cancer-comes-back> [Accessed: 18 July 2019].

80. James ND, Sydes MR, Clarke NW, Mason MD, Dearnaley DP, Anderson J, et al. STAMPEDE: Systemic Therapy for Advancing or Metastatic Prostate Cancer—a multi-arm multi-stage randomised controlled trial. *Clinical Oncology*. (2008)20:577-581

81. Sydes MR, Parmar MKB, Mason MD, Clarke NW, Amos C, Anderson J, et al. Flexible trial design in practice—stopping arms for lack-of-benefit and adding research arms mid-trial in STAMPEDE: a multi-arm multi-stage randomized controlled trial. *Trials*. 2012;13(1):168.

82. Parmar MKB, Barthel FMS, Sydes M, Langley R, Kaplan R, Eisenhauer E, et al. Speeding up the evaluation of new agents in cancer. *Journal of the National Cancer Institute*. 2008;100(17):1204-14.

83. James ND, Sydes MR, Clarke NW, Mason MD, Dearnaley DP, Spears MR, et al. Addition of docetaxel, zoledronic acid, or both to first-line long-term hormone

therapy in prostate cancer (STAMPEDE): survival results from an adaptive, multiarm, multistage, platform randomised controlled trial. *The Lancet*. 2016;387(10024):1163-77.

84. James ND, Sydes MR, Mason MD, Clarke NW, Anderson J, Dearnaley DP, et al. Celecoxib plus hormone therapy versus hormone therapy alone for hormone-sensitive prostate cancer: first results from the STAMPEDE multiarm, multistage, randomised controlled trial. *The Lancet oncology*. 2012;13(5):549-58.

85. James ND, de Bono JS, Spears MR, Clarke NW, Mason MD, Dearnaley DP, et al. Abiraterone for prostate cancer not previously treated with hormone therapy. *New England Journal of Medicine*. 2017;377(4):338-51.

86. Parker CC, James ND, Brawley CD, Clarke NW, Hoyle AP, Ali A, et al. Radiotherapy to the primary tumour for newly diagnosed, metastatic prostate cancer (STAMPEDE): a randomised controlled phase 3 trial. *The Lancet*. 2018;392(10162):2353-66.

87. Hotte SJ, Saad F. Current management of castrate-resistant prostate cancer. *Current oncology*. 2010;17(Suppl 2):S72.

88. Woods BS, Sideris E, Sydes MR, Gannon MR, Parmar MK, Alzouebi M, et al. Addition of docetaxel to first-line long-term hormone therapy in prostate cancer (STAMPEDE): modelling to estimate long-term survival, quality-adjusted survival, and cost-effectiveness. *European urology oncology*. 2018;1(6):449-58.

89. Mahmud SM, Fong B, Fahmy N, Tanguay S, Aprikian AG. Effect of preoperative delay on survival in patients with bladder cancer undergoing cystectomy in Quebec: a population based study. *The Journal of urology*. 2006;175(1):78-83.

90. National Institute for Health and Care Excellence (UK). NICE guideline. No 2 (2005). Bladder cancer: diagnosis and management. 2015.

91. Kulkarni GS, Hakenberg OW, Gschwend JE, Thalmann G, Kassouf W, Kamat A, et al. An updated critical analysis of the treatment strategy for newly diagnosed high-grade T1 (previously T1G3) bladder cancer. *European urology*. 2010;57(1):60-70.

92. Engilbertsson H, Aaltonen KE, Björnsson S, Kristmundsson T, Patschan O, Rydén L, et al. Transurethral bladder tumor resection can cause seeding of cancer cells into the bloodstream. *The Journal of Urology*. 2015;193(1):53-7.

93. Mitchell C, Parikh O. Delays in Treatment of Muscle-invasive Bladder Carcinoma. *Clinical Oncology*. 2014;26(2):e6.

94. Sanchez-Ortiz RF, Huang WC, Mick R, Van Arsdalen KN, Wein AJ, Malkowicz SB. An interval longer than 12 weeks between the diagnosis of muscle

invasion and cystectomy is associated with worse outcome in bladder carcinoma. The Journal of Urology. 2003;169(1):110-5.

95. Cancer waiting times. Cancer Research UK [ONLINE] Available at: <https://www.cancerresearchuk.org/about-cancer/cancer-in-general/treatment/access-to-treatment/waiting-times-after-diagnosis> [Accessed: 05 April 2019].

96. Kulkarni GS, Urbach DR, Austin PC, Fleshner NE, Laupacis A. Longer wait times increase overall mortality in patients with bladder cancer. The Journal of urology. 2009;182(4):1318-24.

97. Fahmy NM, Mahmud S, Aprikian AG. Delay in the surgical treatment of bladder cancer and survival: systematic review of the literature. European urology. 2006;50(6):1176-82.

98. Xie W, Regan M, Buyse M, Halabi S, Kantoff P, Sartor O, et al. Metastasis-free survival is a strong surrogate of overall survival in localized prostate cancer. Journal of Clinical Oncology. 2017;35(27):3097-114.

99. Sydes MR, Mason MD, Spears MR, Clarke NW, Dearnaley D, Ritchie AWS, et al. Adding abiraterone acetate plus prednisolone (AAP) or docetaxel for patients (pts) with high-risk prostate cancer (PCa) starting long-term androgen deprivation therapy (ADT): Directly randomised data from STAMPEDE (NCT00268476). Annals of Oncology. 2017;28(suppl_5).

100. Shaw B, Mansfield FL, Borges L. One-stage posterolateral decompression and stabilization for primary and metastatic vertebral tumors in the thoracic and lumbar spine. Journal of neurosurgery. 1989;70(3):405-10.

101. Maccauro G, Spinelli MS, Mauro S, Perisano C, Graci C, Rosa MA. Physiopathology of spine metastasis. International journal of surgical oncology. 2011;2011:107969

102. Jefferies ER, Bahl A, Hounscome L, Eylert MF, Verne J, Persad RA, et al. Admissions to hospital due to fracture in England in patients with prostate cancer treated with androgen-deprivation therapy—do we have to worry about the hormones? British Journal of Urology (BJU) international. 2016;118(3):416-22.

103. Kilburn LS, Aresu M, Banerji J, Barrett-Lee P, Ellis P, Bliss JM. Can routine data be used to support cancer clinical trials? A historical baseline on which to build: retrospective linkage of data from the TACT (CRUK 01/001) breast cancer trial and the National Cancer Data Repository. Trials. 2017;18(1):561.

104. Franklin M, Thorn J. Self-reported and routinely collected electronic healthcare resource-use data for trial-based economic evaluations: the current state of

play in England and considerations for the future. BMC medical research methodology. 2019;19(1):1-13.

105. Wright-Hughes A, Graham E, Cottrell D, Farrin A. Routine hospital data— is it good enough for trials? An example using England’s Hospital Episode Statistics in the SHIFT trial of Family Therapy vs. Treatment as Usual in adolescents following self-harm. Clinical trials. 2018;15(2):197-206.

106. Röhrig B, du Prel J-B, Wachtlin D, Blettner M. Types of study in medical research. Deutsches Arzteblatt International. 2009;106(15):262-8.

107. Al-Busaidi ZQ. Qualitative research and its uses in health care. Sultan Qaboos University Medical Journal. 2008;8(1):11.

108. Endnote. Clarivate analytics [ONLINE] Available at: <https://www.myendnoteweb.com/EndNoteWeb.html> [Accessed: 15 Aug 2019].

109. R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria [ONLINE] Available at: <https://www.R-project.org/> [Accessed: 05 Sept 2019].

110. RStudio Team (2016). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA [ONLINE] Available at: <http://www.rstudio.com/> [Accessed: 05 Sept 2019].

111. HSCIC changing its name to NHS Digital. Gov.uk [ONLINE] Available at: <https://www.gov.uk/government/news/hscic-changing-its-name-to-nhs-digital> [Accessed: 06 Sept 2019].

112. Last JM, Harris SS, Thuriaux MC, Spasoff RA. A dictionary of epidemiology: International Epidemiological Association, Inc.; 2001.

113. What is a Serious Adverse Event? US Food and Drug Administration (FDA) [ONLINE] Available at: <https://www.fda.gov/safety/medwatch/howtoreport/ucm053087.htm> [Accessed: 11 March 2019].

114. Sydes MR, Ritchie AWS, James ND. Reply to Che-Kai Tsao, Matthew D. Galsky, and William K. Oh's Platinum Opinion. Docetaxel for Metastatic Hormone-Sensitive Prostate Cancer: Urgent Need To Minimize The Risk Of Neutropenic Fever. European urology. (2016):707-708.

115. Markman M. Cross-trial comparisons in the oncology arena: when is this justified? Oncology. 2011;80(3-4):151-2.

116. Clarke RT, Jenyon T, van Hamel Parsons V, King AJ. Neutropenic sepsis: management and complications. *Clinical Medicine*. 2013;13(2):185-7.
117. Penack O, Buchheidt D, Christopeit M, von Lilienfeld-Toal M, Massenkeil G, Hentrich M, et al. Management of sepsis in neutropenic patients: guidelines from the infectious diseases working party of the German Society of Hematology and Oncology. *Annals of oncology*. 2010;22(5):1019-29.
118. Dellinger RP, Levy MM, Carlet JM, Bion J, Parker MM, Jaeschke R, et al. Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock: 2008. *Intensive care medicine*. 2008;34(1):17-60.
119. Kuderer NM, Dale DC, Crawford J, Cosler LE, Lyman GH. Mortality, morbidity, and cost associated with febrile neutropenia in adult cancer patients. *Cancer*. 2006;106(10):2258-66.
120. Talbot T, Dangoor A, Shah R, Naik J, Talbot D, Lester JF, et al. The burden of neutropenic sepsis in patients with advanced non-small cell lung cancer treated with single-agent docetaxel: A retrospective study. *Lung Cancer*. 2017;113:115-20.
121. Neutropenic sepsis: Prevention and management in people with cancer. National Institute for Health and Care Excellence (NICE) [ONLINE] Available at: <https://www.nice.org.uk/guidance/cg151/chapter/1-Guidance#managing-suspected-neutropenic-sepsis-in-secondary-and-tertiary-care-2> [Accessed: 29 Nov 2018].
122. National Institute for Health and Care Excellence (UK). NICE guideline. Neutropenic sepsis: Prevention and management of neutropenic sepsis in cancer patients. 2012.
123. Moore DC. Drug-induced neutropenia: a focus on rituximab-induced late-onset neutropenia. *Pharmacy and therapeutics*. 2016;41(12):765.
124. Coding of sepsis. NHS Digital [ONLINE] Available at: https://hscic.kahootz.com/connect.ti/t_c_home/viewQuestionnaire?qid=4638467 [Accessed: 13 Dec 2018].
125. Consultation Results for the Proposed Update to DChS.I.1 Sepsis. NHS Digital [ONLINE] Available at: https://hscic.kahootz.com/connect.ti/t_c_home/view?objectId=389060#389060 [Accessed: 05 Sept 19].
126. Sweeney CJ, Chen Y-H, Carducci M, Liu G, Jarrard DF, Eisenberger M, et al. Chemohormonal therapy in metastatic hormone-sensitive prostate cancer. *New England Journal of Medicine*. 2015;373(8):737-46.

127. Tucci M, Bertaglia V, Vignani F, Buttigliero C, Fiori C, Porpiglia F, et al. Addition of docetaxel to androgen deprivation therapy for patients with hormone-sensitive metastatic prostate cancer: a systematic review and meta-analysis. *European urology*. 2016;69(4):563-73.
128. Lalami Y, Paesmans M, Muanza F, Barette M, Plehiers B, Dubreucq L, et al. Can we predict the duration of chemotherapy-induced neutropenia in febrile neutropenic patients, focusing on regimen-specific risk factors? A retrospective analysis. *Annals of oncology*. 2005;17(3):507-14.
129. Gravis G, Fizazi K, Joly F, Oudard S, Priou F, Esterni B, et al. Androgen-deprivation therapy alone or with docetaxel in non-castrate metastatic prostate cancer (GETUG-AFU 15): a randomised, open-label, phase 3 trial. *The lancet oncology*. 2013;14(2):149-58.
130. Tannock IF, de Wit R, Berry WR, Horti J, Pluzanska A, Chi KN, et al. Docetaxel plus prednisone or mitoxantrone plus prednisone for advanced prostate cancer. *New England Journal of Medicine*. 2004;351(15):1502-12.
131. ICH Harmonised Tripartite Guideline. Clinical safety data management: definitions and standards for expedited reporting E2A. International conference on harmonisation of technical requirements for registration of pharmaceuticals for human use. 1994.
132. National Cancer Institute. Common Terminology Criteria for Adverse Events (CTCAE) [ONLINE] Available at: https://ctep.cancer.gov/protocoldevelopment/electronic_applications/ctc.htm#ctc_50 [Accessed: 09 Jan 2019].
133. Mahil J, Hughes C, Patel K, Lyons J, Elliott P, Choudhury A, et al. Febrile neutropenia rates in men treated with docetaxel chemotherapy for metastatic hormone-sensitive prostate cancer. *Clinical Oncology*. 2016;28(9):612.
134. Tanguay JS, Catlow C, Smith C, Barber J, Staffurth J, Kumar S, et al. High risk of neutropenia for hormone-naïve prostate cancer patients receiving STAMPEDE-style upfront docetaxel chemotherapy in usual clinical practice. *Clinical Oncology*. 2016;28(9):611.
135. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine*. 2009;151(4):264-9.
136. Higgins JP, Green S. *Cochrane handbook for systematic reviews of interventions*: John Wiley & Sons; 2011.

137. Google Scholar. Google [ONLINE] Available at: <https://scholar.google.com> [Accessed: 22 Aug 19].
138. Schweizer MT, Gulati R, Mostaghel EA, Nelson PS, Montgomery RB, Evan YY, et al. Docetaxel-related toxicity in metastatic hormone-sensitive and metastatic castration-resistant prostate cancer. *Medical Oncology*. 2016;33(7):77.
139. Maria BA, Merida Garcia A, Lora D, Bernard BD, David L, Manneh Kopp R, et al. Association of risk of febrile neutropenia (FN) with docetaxel in prostate cancer (PC) patients: A meta-analysis of published phase II-III trials. *American Society of Clinical Oncology*. 2017;35(15 supplement e21683).
140. Tsao CK, Galsky MD, Oh WK. Docetaxel for Metastatic Hormone-sensitive Prostate Cancer: Urgent Need to Minimize the Risk of Neutropenic Fever. *European Urology*. 2016 Nov;70(5):707-708.
141. James ND, Sydes MR, Clarke NW, Parmar MKB, Ritchie A. Response to 'High Risk of Neutropenia for Hormone-naïve Prostate Cancer Patients Receiving STAMPEDE-style Upfront Docetaxel Chemotherapy in Usual Clinical Practice', by Tanguay et al. *Clinical Oncology*. 2016;28(10):666-7.
142. Templeton A, Vera-Badillo F, Wang L, Attalla M, De Gouveia P, Leibowitz-Amit R, et al. Translating clinical trials to clinical practice: outcomes of men with metastatic castration resistant prostate cancer treated with docetaxel and prednisone in and out of clinical trials. *Annals of oncology*. 2013;24(12):2972-7.
143. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of chronic diseases*. 1987;40(5):373-83.
144. Lee CK, Lord SJ, Stockler MR, Coates AS, Gebiski V, Simes RJ. Historical cross-trial comparisons for competing treatments in advanced breast cancer—an empirical analysis of bias. *European Journal of Cancer*. 2010;46(3):541-8.
145. Saad F, Hotte SJ. Guidelines for the management of castrate-resistant prostate cancer. *Canadian Urological Association Journal*. 2010;4(6):380.
146. Guo B, Yuan Y. A comparative review of methods for comparing means using partially paired data. *Statistical methods in medical research*. 2017;26(3):1323-40.
147. Claassen JAHR. The gold standard: not a golden standard. *British Medical Journal*. 2005;330(7500):1121.
148. Ibrahim JG, Chu H, Chen MH. Missing data in clinical studies: issues and methods. *Journal of clinical oncology*. 2012;30(26):3297.

149. Kaushal S. Missing data in clinical trials: Pitfalls and remedies. *International journal of applied & basic medical research*. 2014;4(Suppl 1):S6-7.
150. Muhammad IA. Febrile neutropenia rates in men treated with Docetaxel chemotherapy for metastatic hormone sensitive prostate cancer [UNPUBLISHED MSc thesis]. 2018.
151. The ICECaP Working Group. The development of intermediate clinical endpoints in cancer of the prostate (ICECaP). *Journal of the National Cancer Institute*. 2015;107(12):djv261.
152. Wilson R. Patient led PROMs must take centre stage in cancer research. *Research Involvement and Engagement*. 2018;4(1):7.
153. Ulrich CM, Knafl KA, Ratcliffe SJ, Richmond TS, Grady C, Miller-Davis C, et al. Developing a model of the benefits and burdens of research participation in cancer clinical trials. *AJOB primary research*. 2012;3(2):10-23.
154. Fogel DB. Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: a review. *Contemporary clinical trials communications*. 2018;11:156-64.
155. Kristman V, Manno M, Côté P. Loss to follow-up in cohort studies: how much is too much? *European journal of epidemiology*. 2004;19(8):751-60.
156. Scott R, Group MASS. The Multicentre Aneurysm Screening Study (MASS) into the effect of abdominal aortic aneurysm screening on mortality in men: a randomised controlled trial. *The Lancet*. 2002;360(9345):1531-9.
157. Web of Science. Clarivate analytics [ONLINE] Available at: https://apps.webofknowledge.com/WOS_GeneralSearch_input.do?product=WOS&search_mode=GeneralSearch&SID=C4duPbl7h5bjvwz7dxX&preferencesSaved= [Accessed: 15 Aug 2019].
158. PubMed. National Centre for Biotechnology Information (NCBI) [ONLINE] Available at: <http://www.pubmed.gov> [Accessed: 15 Aug 2019].
159. MEDLINE. Ovid [ONLINE] Available at: <https://ovidsp.dc1.ovid.com/sp-4.01.0a/ovidweb.cgi> [Accessed: 15 Aug 2019].
160. Critical appraisal skills programme (CASP) [ONLINE] Available at: <https://casp-uk.net> [Accessed: 30 July 2019].

161. Snilstveit B, Oliver S, Vojtkova M. Narrative approaches to systematic review and synthesis of evidence for international development policy and practice. *Journal of development effectiveness*. 2012;4(3):409-29.
162. Fletcher J. What is heterogeneity and is it important? *Biomedical Journal*. 2007;334(7584):94-6.
163. Earle CC, Nattinger AB, Potosky AL, Lang K, Mallick R, Berger M, et al. Identifying cancer relapse using SEER-Medicare data. *Medical Care*. 2002;40(8 Suppl):IV-75-81.
164. McClish D, Penberthy L, Pugh A. Using Medicare claims to identify second primary cancers and recurrences in order to supplement a cancer registry. *Journal of clinical epidemiology*. 2003;56(8):760-7.
165. Lamont EB, Herndon JE, 2nd, Weeks JC, Henderson IC, Earle CC, Schilsky RL, et al. Measuring disease-free survival and cancer relapse using Medicare claims from CALGB breast cancer trial participants (companion to 9344). *Journal of the National Cancer Institute*. 2006;98(18):1335-8.
166. Anaya DA, Becker NS, Richardson P, Abraham NS. Use of administrative data to identify colorectal liver metastasis. *Journal of Surgical Research*. 2012;176(1):141-6.
167. Chubak J, Yu O, Pocobelli G, Lamerato L, Webster J, Prout MN, et al. Administrative data algorithms to identify second breast cancer events following early-stage invasive breast cancer. *Journal of the National Cancer Institute*. 2012;104(12):931-40.
168. Kimmick GG, Camacho F, Hwang W, Mackley H, Stewart J, Anderson RT. Adjuvant radiation and outcomes after breast conserving surgery in publicly insured patients. *Journal of geriatric oncology*. 2012;3(2):138-46.
169. Nordstrom BL, Whyte JL, Stolar M, Mercaldi C, Kallich JD. Identification of metastatic cancer in claims data. *Pharmacoepidemiology and Drug Safety*. 2012;21(SUPPL.2):21-8.
170. Hagberg KW, Taylor A, Hernandez RK, Jick S. Incidence of bone metastases in breast cancer patients in the United Kingdom: Results of a multi-database linkage study using the general practice research database. *Cancer Epidemiology*. 2013;37(3):240-6.
171. Lash TL, Riis AH, Ostfeld EB, Erichsen R, Vyberg M, Thorlacius-Ussing O. A validated algorithm to ascertain colorectal cancer recurrence using registry resources in Denmark. *International Journal of Cancer*. 2015;136(9):2210-5.

172. Ricketts K, Williams M, Liu ZW, Gibson A. Automated estimation of disease recurrence in head and neck cancer using routine healthcare data. *Computer Methods and Programs in Biomedicine*. 2014;117(3):412-24.
173. Ehrenstein V, Hernandez RK, Maegbaek ML, Kahlert J, Nguyen-Nielsen M, Norgaard M, et al. Validation of algorithms to detect distant metastases in men with prostate cancer using routine registry data in Denmark. *Clinical Epidemiology*. 2015;7(pp 259-265).
174. Deshpande AD, Schootman M, Mayer A. Development of a claims-based algorithm to identify colorectal cancer recurrence. *Annals of Epidemiology*. 2015;25(4):297-300.
175. Haque R, Shi J, Schottinger JE, Ahmed SA, Chung J, Avila C, et al. A hybrid approach to identify subsequent breast cancer using pathology and automated health information data. *Medical Care*. 2015;53(4):380-5.
176. Warren JL, Mariotto A, Melbert D, Schrag D, Doria-Rose P, Penson D, et al. Sensitivity of Medicare Claims to Identify Cancer Recurrence in Elderly Colorectal and Breast Cancer Patients. *Medical care*. 2016;54(8):e47-54.
177. Livaudais-Toman J, Egorova N, Franco R, Prasad-Hayes M, Howell EA, Wisnivesky J, et al. A Validation Study of Administrative Claims Data to Measure Ovarian Cancer Recurrence and Secondary Debulking Surgery. *The Journal for Electronic Data and Methods (eGEMs)*. 2016;4(1):1208.
178. Nordstrom BL, Simeone JC, Malley KG, Fraeman KH, Klippel Z, Durst M, et al. Validation of claims algorithms for progression to metastatic cancer in patients with breast, non-small cell lung, and colorectal cancer. *Frontiers in Oncology*. 2016;6:18.
179. Joshi V, Adelstein BA, Schaffer A, Srasuebkul P, Dobbins T, Pearson SA. Validating a proxy for disease progression in metastatic cancer patients using prescribing and dispensing data. *Asia Pacific Journal of Clinical Oncology*. 2017;13(5):e246-52.
180. Hassett MJ, Uno H, Cronin AM, Carroll NM, Hornbrook MC, Ritzwoller D. Detecting Lung and Colorectal Cancer Recurrence Using Structured Clinical/Administrative Data to Enable Outcomes Research and Population Health Management. *Medical Care*. 2017 Dec;55(12):e88.
181. Gupta S, Nathan PC, Baxter NN, Lau C, Daly C, Pole JD. Validity of Administrative Data in Identifying Cancer-related Events in Adolescents and Young Adults: A Population-based Study Using the IMPACT Cohort. *Medical Care*. 2018;56(6):e32-8.
182. Ritzwoller DP, Hassett MJ, Uno H, Cronin AM, Carroll NM, Hornbrook MC, et al. Development, validation, and dissemination of a breast cancer recurrence

detection and timing informatics algorithm. *Journal of the National Cancer Institute*. 2017;110(3):273-81.

183. Wong S-L, Ricketts K, Royle G, Williams M, Mendes R. A methodology to extract outcomes from routine healthcare data for patients with locally advanced non-small cell lung cancer. *BioMed Central health services research*. 2018;18(1):278.

184. Rasmussen LA, Jensen H, Virgilsen LF, Jensen JB, Vedsted P. A validated algorithm to identify recurrence of bladder cancer: a register-based study in Denmark. *Clinical epidemiology*. 2018;10:1755.

185. Uno H, Ritzwoller DP, Cronin AM, Carroll NM, Hornbrook MC, Hassett MJ. Determining the time of cancer recurrence using claims or electronic medical record data. *Journal of Clinical Oncology (JCO) clinical cancer informatics*. 2018;2:1-10.

186. Xu Y, Kong S, Cheung WY, Quan ML, Nakoneshny SC, Dort JC. Developing case-finding algorithms for second events of oropharyngeal cancer using administrative data: A population-based validation study. *Head & Neck*. 2019;41(7):2291-2298.

187. Xu Y, Kong S, Cheung WY, Bouchard-Fortier A, Dort JC, Quan H, et al. Development and validation of case-finding algorithms for recurrence of breast cancer using routinely collected administrative data. *BioMed central cancer*. 2019;19(1):210.

188. Rasmussen LA, Jensen H, Virgilsen LF, Thorsen LBJ, Offersen BV, Vedsted P. A validated algorithm for register-based identification of patients with recurrence of breast cancer—Based on Danish Breast Cancer Group (DBCG) data. *Cancer epidemiology*. 2019;59:129-34.

189. Warren JL, Yabroff KR. Challenges and Opportunities in Measuring Cancer Recurrence in the United States. *Journal of the National Cancer Institute*. 2015;107(8).

190. Sim J, Lewis M. The size of a pilot study for a clinical trial should be calculated in relation to considerations of precision and efficiency. *Journal of clinical epidemiology*. 2012;65(3):301-8.

191. Tse T, Williams RJ, Zarin DA. Reporting “basic results” in ClinicalTrials.gov. *Chest*. 2009;136(1):295-303.

192. Schmidt M, Schmidt SAJ, Sandegaard JL, Ehrenstein V, Pedersen L, Sørensen HT. The Danish National Patient Registry: a review of content, data quality, and research potential. *Clinical epidemiology*. 2015;7:449.

193. Sharma M, Sarin A, Gupta P, Sachdeva S, Desai AV. Journal impact factor: its use, significance and limitations. *World journal of nuclear medicine*. 2014;13(2):146.

194. Huh S. How to add a journal to the international databases, Science Citation Index Expanded and MEDLINE. Archives of plastic surgery. 2016;43(6):487.
195. Buyse M, Molenberghs G. Criteria for the validation of surrogate endpoints in randomized experiments. Biometrics. 1998:1014-29.
196. Buyse M, Molenberghs G, Burzykowski T, Renard D, Geys H. The validation of surrogate endpoints in meta-analyses of randomized experiments. Biostatistics. 2000;1(1):49-67.
197. Prostate cancer: diagnosis and management. National Institute for Health and Care Excellence (NICE) [ONLINE] Available at: <https://www.nice.org.uk/guidance/cg175> [Accessed: 21 Feb 2018].
198. Systemic therapy in Advancing or Metastatic Prostate Cancer: Evaluation of Drug Efficacy, A multi-arm multi-stage randomised controlled trial. The protocol v17 [ONLINE] Available at: http://www.stampededtrial.org/media/1841/01-stampede_protocol_v17_clean.pdf [Accessed: 05 Sept 2019].
199. Hinkle DE, Wiersma W, Jurs SG. Applied statistics for the behavioral sciences: Houghton Mifflin College Division; 2003.
200. Mukaka MM. A guide to appropriate use of correlation coefficient in medical research. Malawi Medical Journal. 2012;24(3):69-71.
201. McCowan C, Thomson E, Szmigielski CA, Kalra D, Sullivan FM, Prokosch H-U, et al. Using electronic health records to support clinical trials: a report on stakeholder engagement for EHR4CR. BioMed research international. 2015.
202. Oxnard GR, Morris MJ, Hodi FS, Baker LH, Kris MG, Venook AP, et al. When progressive disease does not mean treatment failure: reconsidering the criteria for progression. Journal of the National Cancer Institute. 2012;104(20):1534-41.
203. Kemp R, Prasad V. Surrogate endpoints in oncology: when are they acceptable for regulatory and clinical decisions, and are they currently overused? BioMed central medicine. 2017;15(1):134.
204. Sarkar S, Seshadri D. Conducting record review studies in clinical practice. Journal of clinical and diagnostic research: Journal of Clinical and Diagnostic Research (JCDR). 2014;8(9):JG01.
205. NHS Digital data and information strategy. NHS Digital [ONLINE] Available at: <https://digital.nhs.uk/data-and-information/strategy> [Accessed: 13 Feb 2018].

206. The processing cycle and HES data quality. NHS Digital [ONLINE] Available at: <http://content.digital.nhs.uk/article/1825/The-processing-cycle-and-HES-data-quality> [Accessed 13 Feb. 2018]
207. Panageas KS, Ben-Porat L, Dickler MN, Chapman PB, Schrag D. When you look matters: the effect of assessment schedule on progression-free survival. *Journal of the National Cancer Institute*. 2007;99(6):428-32.
208. Grann AF, Erichsen R, Nielsen AG, Frøslev T, Thomsen RW. Existing data sources for clinical epidemiology: the clinical laboratory information system (LABKA) research database at Aarhus University, Denmark. *Clinical epidemiology*. 2011;3:133.
209. Wood C, Lawton S, Downing A, Morris E, Glaser AW, Thomas J, et al. Quality of life of colorectal cancer survivors in England: report on a national survey of colorectal cancer survivors using Patient Reported Outcome Measures (PROMs). NHS England. 2015.
210. Glaser AW, Fraser LK, Corner J, Feltbower R, Morris EJ, Hartwell G, et al. Patient-reported outcomes of cancer survivors in England 1–5 years after diagnosis: a cross-sectional survey. *Biomedical journal (BMJ) open*. 2013;3(4):e002317.
211. Bland JM, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *The lancet*. 1986;327(8476):307-10.
212. Risbridger GP, Davis ID, Birrell SN, Tilley WD. Breast and prostate cancer: more similar than different. *Nature Reviews Cancer*. 2010;10(3):205.
213. Rosa C, Campbell AN, Miele GM, Brunner M, Winstanley EL. Using e-technologies in clinical trials. *Contemporary clinical trials*. 2015;45:41-54.
214. Lo B. Sharing clinical trial data: maximizing benefits, minimizing risk. *Jama*. 2015;313(8):793-4.
215. O’Leary E, Seow H, Julian J, Levine M, Pond GR. Data collection in cancer clinical trials: Too much of a good thing? *Clinical Trials*. 2013;10(4):624-32.
216. Williams JG, Cheung WY, Cohen DR, Hutchings HA, Longo MF, Russell IT. Can randomised trials rely on existing electronic data? A feasibility study to explore the value of routine data in health technology assessment. *Health technology assessment (Winchester, England)*. 2003;7(26):iii-v.
217. Raftery J, Roderick P, Stevens A. Potential use of routine databases in health technology assessment. *Health technology assessment*. 2005;9(20):1-106.

218. Bhattacharya I, Morden J, Griffin C, Snowdon C, Brannan R, Bliss J, et al. The application and feasibility of using routine data sources for long-term cancer clinical trial follow-up. *Clinical oncology*. 2017;29(12):796.
219. Powell GA, Bonnett LJ, Tudur-Smith C, Hughes DA, Williamson PR, Marson AG. Using routinely recorded data in the UK to assess outcomes in a randomised controlled trial: The Trials of Access. *Trials*. 2017;18(1):389.
220. Thorn JC, Turner EL, Hounscome L, Walsh E, Down L, Verne J, et al. Validating the use of Hospital Episode Statistics data and comparison of costing methodologies for economic evaluation: an end-of-life case study from the Cluster randomised triAl of PSA testing for Prostate cancer (CAP). *Biomedical Journal (BMJ) open*. 2016;6(4):e011063.
221. Murray DW, MacLennan GS, Breeman S, Dakin HA, Johnston L, Campbell MK, et al. A randomised controlled trial of the clinical effectiveness and cost-effectiveness of different knee prostheses: the Knee Arthroplasty Trial (KAT). *Health technology assessment*. 2014;18(19):1.
222. Van Staa TP, Goldacre B, Gulliford M, Cassell J, Pirmohamed M, Taweel A, et al. Pragmatic randomised trials using routine electronic health records: putting them to the test. *Bmj*. 2012;344:e55.
223. Van Staa TP, Dyson L, McCann G, Padmanabhan S, Belatri R, Goldacre B, et al. The opportunities and challenges of pragmatic point-of-care randomised trials using routinely collected electronic records: evaluations of two exemplar trials. *Health technology assessment*. 2014;18(43):1-146.
224. Van Staa TP, Goldacre B, Gulliford M, Cassell J, Pirmohamed M, Taweel A, et al. Randomised Evaluations of Accepted Choices in Treatment (REACT) trials: large-scale pragmatic trials within databases of routinely collected electronic healthcare records. *Trials*. 2011;12(1):A104.
225. National Institute for Health Research (NIHR). PRIMETIME – Post-operative avoidance of radiotherapy in minimal risk women: patient selection using biomarkers [ONLINE] Available at: <https://ukctg.nihr.ac.uk/trials/trial-details/trial-details?trialNumber=ISRCTN41579286> [Accessed: 05 April 2019].
226. Gulliford MC, van Staa T, McDermott L, Dregan A, McCann G, Ashworth M, et al. Cluster randomised trial in the General Practice Research Database: 1. Electronic decision support to reduce antibiotic prescribing in primary care (eCRT study). *Trials*. 2011;12(1):115.
227. Dregan A, van Staa T, McDermott L, McCann G, Ashworth M, Charlton J, et al. Cluster randomized trial in the general practice research database: 2. Secondary

prevention after first stroke (eCRT study): study protocol for a randomized controlled trial. *Trials*. 2012;13(1):181.

228. Gulliford MC, van Staa TP, McDermott L, McCann G, Charlton J, Dregan A. Cluster randomized trials utilizing primary care electronic health records: methodological issues in design, conduct, and analysis (eCRT Study). *Trials*. 2014;15(1):220.

229. Data and Audit Project. The British Association of Urological Surgeons [ONLINE] Available at: https://www.baus.org.uk/professionals/baus_business/data_audit.aspx [Accessed: 19 Oct 2018].

230. Jefferies ER, Cresswell J, McGrath JS, Miller C, Hounsborne L, Fowler S, et al. Open radical cystectomy in England: the current standard of care—an analysis of the British Association of Urological Surgeons (BAUS) cystectomy audit and Hospital Episodes Statistics (HES) data. *British Journal of Urology (BJU) international*. 2018;121(6):880-5.

231. Main Speciality Code. NHS Digital [ONLINE] Available at: https://www.datadictionary.nhs.uk/data_dictionary/attributes/m/main_specialty_code_de.asp [Accessed: 22 Aug 2018].

232. Payment by results. NHS Digital [ONLINE] Available at: <https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics/payment-by-results> [Accessed: 02 July 2019].

233. The quality of the clinical coding in the NHS. CAPITA [ONLINE] Available at: http://www.chks.co.uk/userfiles/files/The_quality_of_clinical_coding_in_the_NHS.pdf [Accessed: 02 July 2019].

234. Spine. NHS Digital [ONLINE] Available at: <https://digital.nhs.uk/services/spine> [Accessed: 14 Nov 2018].

235. National Radiotherapy Database. National Cancer Registration and Analysis Service [ONLINE] Available at: http://www.ncin.org.uk/collecting_and_using_data/rtds [Accessed: 23 Aug 2018].

236. Diagnostic Imaging Dataset. NHS England [ONLINE] Available at: <https://www.england.nhs.uk/statistics/statistical-work-areas/diagnostic-imaging-dataset/> [Accessed: 23 Aug 2018].

237. Little RJ, D'Agostino R, Cohen ML, Dickersin K, Emerson SS, Farrar JT, et al. The prevention and treatment of missing data in clinical trials. *New England Journal of Medicine*. 2012;367(14):1355-60.

238. Data access request service (DARS) charges 2018/19. NHS Digital [ONLINE] Available at: <https://digital.nhs.uk/services/data-access-request-service-dars/data-access-request-service-dars-charges-2018-19> [Accessed: 10 July 19].

8 Appendices

8.1 Chapter three

8.1.1 PRISMA guidelines (neutropenic events review)

Section/topic	#	Checklist item	Reported?
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	✓
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	✗
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	✓
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	✓
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	✗
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	✓
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	✓
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	✓
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	✓
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	✓
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	✓
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	✓
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	✓
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	✗
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	✓
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	✓

Section/topic	#	Checklist item	Reported?
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	✓
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	✓
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	✓ (discussion)
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	✓
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	✗
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	✓ (discussion)
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	✓ (discussion)
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	✓
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	✓
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	✓
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	✓ (section funding)

Figure 45: The PRISMA guidelines used to undertake the systematic review.

8.1.2 Note review confirmed inferred events

Neutropenic event (HES inferral)	Found without inferral?	Reason event missed without inferral	Chemotherapy identifiable in SACT?
1	✗	Chemotherapy HES data missing	Prior to SACT collection
2	✗	HES event 12 weeks after last HES chemotherapy	-
3	✓	-	-
4	✗	<i>Two instances</i> ; HES event 18 weeks after and 102 weeks before last HES chemotherapy	-
5	✓	-	-
6	✓	-	-
7	✓	-	-
8	✓	-	-
9	✗	Chemotherapy HES data missing	✓
10	✗	Chemotherapy HES data missing	Prior to SACT collection
11	✓	-	-
12	✗	Chemotherapy HES data missing	Prior to SACT collection
13	✗	HES event 40 weeks after last HES chemotherapy	-
14	✓	-	-
15	✓	-	-
16	✓	-	-
17	✗	Chemotherapy HES data missing	✓
18	✗	Chemotherapy HES data missing	✓
19*	✓*	-	-
20*	✓*	-	-
21*	✓*	-	-
22*	✓*	-	-

Table 78: The 22 patients that experienced a neutropenic event identified using the HES algorithm with inferral. Whether or not the algorithm could identify these events without inferral is also present. The reason for the missing data is also described, including whether the chemotherapy data could be identified using the SACT data.

* = Occurring at HSPC. If not specified, then the event occurred at CRPC; ✗ = missed; ✓ = identified.

8.1.3 PHE application

Section 3: project summary (mandatory)

All fields in this section must be completed. See guidance for more information.

Applicants are advised that all requests to access PHE data must be accompanied by a detailed protocol.

Project overview	
ODR reference:	Insert any ODR reference assigned for this project. Leave blank if new project.
Data sharing contract reference:	ODR1718_094 - Warwick Med Sch & Queen Elizabeth Hospital - Hattie Mintz (STAMPEDE) Reference any data sharing contract held with PHE for this project. Leave blank if new project.
Project title:	Using routine data to identify clinical trial outcomes within the STAMPEDE trial
Lay summary	
Describe in plain English the overall project aim(s) and objectives:	Project Aims i. To identify better ways of obtaining clinical trial outcome data to: (A) Repeat reported STAMPEDE analyses to validate the PhD project's algorithm capabilities and the trial results. (B) Perform new, secondary analyses not possible with conventionally-collected trial data, including but not limited to, rates of neutropenic sepsis when chemotherapy is given at different times in their disease or cardiac events on hormone therapy. ii. To develop a clinically useable tool (e.g. algorithm), for use in a clinical trial, to accurately identify disease driven events and trial outcomes. iii. To help reduce the burden of collecting clinical trial data from traditional patient and clinician contact. By using data that has already been accurately collected by the NHS, it may be possible to improve timeliness, reduce costs and save resources that can then be used elsewhere. Objectives <ul style="list-style-type: none">Develop, enhance, and validate methodology to calculate trial outcomes from EHRs. Can the new tool accurately detect disease-related events and trial outcome events, and therefore successfully identify treatment effect differences? The tool must:<ul style="list-style-type: none">Either be able to use HES data alone or use non-HES-based data in addition (i.e. RTDS/SACT, cancer registration)Be clinically useable for planned applicationsIdentify if the routine data can be used directly for clinical trial follow-up and if yes, then utilise this data for routine trial follow-up (potentially to replace active in-trial follow-up completely)

Describe in plain English the rationale for the conduct of the project:	<p>This project aims to show if routine Electronic Healthcare Records (EHRs) can be used to collect data for clinical trials. We believe and have shown [1] the information found in an EHR could enhance clinical trial data collection. For example, as EHRs contain many of the details of a person's health until death, with permission from the patients, researchers can then follow what happens to the patient. Any long term benefits (or harms) can then be explored. We propose that the use of this data could replace routine hospital patient trial follow-up visits completely.</p> <p>This project plans to explore the use of EHRs within the ongoing clinical trial protocol, STAMPEDE (Systemic Therapy for Advancing or Metastatic Prostate cancer: Evaluation of Drug Efficacy). This trial is the largest study of treatments for prostate cancer in the world. Most STAMPEDE patients are from England (85%), with others from Wales (6%), Scotland (7%), N Ireland (2%) and a small number from overseas (1%). To take part in STAMPEDE, patients agree for us to collect long-term survival information, which can be accessed from routine sources of data, such as the Hospital Episode Statistics (HES). The HES is a database of all hospital admissions, outpatient appointments and A&E admissions that happen in NHS Trusts in England.</p> <p>[1] 10.1200/JCO.2017.35.6_suppl.257 Journal of Clinical Oncology 35, no. 6, suppl (February 2017) 257-257.</p>
Describe in plain English how PHE data will be used in the delivery of the project:	<p>To assess the effect of the STAMPEDE treatments from EHRs, trial outcomes must first be calculated. New methods need to be created to be able to quickly capture useful details on clinical trial related events from the EHRs (e.g. progression of the cancer (disease gets worse), or neutropenic sepsis events (complication of cancer treatment, when a patient catches a bacterial infection). Currently, we are developing new ways, using a pilot dataset, but they need to be tested on the large HES dataset. We believe that also using the National Radiotherapy dataset (RTDS), the Systemic Anti-Cancer Therapy dataset (SACT) and the cancer registration will allow a more accurate description of the trial outcomes to be calculated. We aim to test several different methods to identify the clinical trial outcomes to calculate which one works best. These methods include: <i>directly</i> using data fields e.g. using diagnosis and procedure codes to identify sepsis events; and <i>indirectly</i> by creating new surrogate outcomes e.g. to identify progression of the cancer.</p>

<p>Methods</p> <ul style="list-style-type: none"> HES, RTDS, SACT and cancer registration data will be linked to clinical trial number (by PHE) for each consented for flagging in England Methods to detect proposed surrogate outcomes will be applied to the data (HES alone and HES in combination with other routine datasets) Accuracy of RTDS and SACT data will be assessed by comparison to a selected sample of ~50 patients that have been reviewed in depth by clinical review of hospital notes Further validation of RTDS, SACT, HES and cancer registration data will be undertaken with regards to clinical trial use Analysis of events after the allocated treatment in STAMPEDE will be undertaken in relation to treatment arms <p>Anticipated outputs</p> <ul style="list-style-type: none"> <i>Reproduce analysis already carried out within the trial</i> – expect to output Skeletal Related Event data (using the RTDS) on different trial arms <i>Secondary analyses which may not have been possible with conventionally collected trial data</i> e.g. calculate rates of neutropenic sepsis for different patient groups (e.g. different trial arms, progression outcomes etc.) 	<p>Describe in plain English the anticipated public health benefit(s) and/or impact of conducting this project:</p> <p>This analysis will reveal if the use of these datasets could enhance or even replace clinical trial traditional clinic follow-up. An EHR utilising trial system would then be set up for the STAMPEDE trial. The use of EHRs could reduce patient and clinical burden, reduce trial resources and costs and collect more events not routinely collected in the trial.</p>
	<p>Summary of end-use</p>
<p>See information here (hyperlink) for defining research and other types of projects.</p>	<p>Research <input type="checkbox"/> Y</p> <p>Service evaluation <input type="checkbox"/></p> <p>Clinical audit <input type="checkbox"/></p> <p>Surveillance <input type="checkbox"/></p> <p>Other, please specify:</p>

<p>Project timeline</p>	
<p>Please indicate the estimated project start date and project duration (months):</p>	<p>Start date = current</p> <p>End date of PhD (Oct 2019) If we can show that the HES and PHE chemotherapy and radiotherapy records are sufficiently accurate, we would aim to use this as the primary method of collecting that data in an ongoing fashion for the clinical trial.</p>
<p>Section 4: data specification (mandatory)</p> <p>All fields in this section must be completed. See guidance for more information.</p>	
<p>Data specification:</p>	
<p>Classification of data requested (please select appropriate classification):</p> <p>The data is stripped of direct identifiers and techniques such as suppression, offsetting and aggregation are applied to render the data anonymous in line with the ISB Anonymisation Standard for Publishing Health and Social Care Data. The residual risk of re-identification is negligible or very low. Where possible, data will be released under an Open Government Licence with no further control.</p>	<p>Anonymised</p> <p><input type="checkbox"/></p>
<p>The data is stripped of direct identifiers but contains fields which could be used to indirectly identify an individual through combinations of information, either by the people handling the data or by those who see published results (eg ethnicity, sex, month and year of birth, admission dates, geographies or other personal characteristic). The data will be released with controls in line with the ICO Anonymisation Code of Practice.</p>	<p>De-personalised</p> <p><input checked="" type="checkbox"/></p>
<p>The data request includes direct identifiers (eg name, address, NHS number, date of birth) or is coded (pseudonymised), but would be directly identifiable in the hands of the data recipient (such as by hospital number or a cohort-specific identifier). To access identifiable data, an extant legal gateway must be presented (see Section 6). Data will be released with controls.</p>	<p>Personally identifiable</p> <p><input type="checkbox"/></p>
<p>Please indicate the dataset(s) needed for the processing activities (all datasets necessary for the conduct of the project):</p> <p>NCRAS – bespoke extract (RTDS, SACT and cancer registration data)</p> <p>Hospital Episode Statistics (HES) – Admitted care, Accident and Emergency, outpatient</p>	

Specify any data linkage requirements including the required data flows between PHE and the other organisations to be involved. (Where there are multiple data linkages required, involving two or more data processors, the protocol must include a diagram to illustrate the proposed data flows. Please ensure each organisational boundary is clearly identified and where data is moving between organisations, those fields are also included. See guidance for more information and an example data flow diagram.)

• HES (1st pt req, 2005, until most recent
 • data PHE can provide
 • Cancer registration data (-current)
 • RTDS (variables attached) (-current)
 • SACT (variables attached) (-current)

PHE proposed data flow

• Personally-identifiable data sent to PHE from the UCL MRC CTU to allow correct trial patients to be identified and linked to trial number by PHE (e.g. NHS number, DOB)

• PHE linked RTDS, SACT & HES data sent to the UCL MRC CTU (pseudonymised due to stripped NHS number/DOB, identifiable by trial ID)

Analysis

OUTPUT: Assessment of whether the use of EHRs is feasible to replace some aspects of trial follow-up. A method to utilise these EHRs would be developed to allow incorporation into the trial database. Continual feed of data from routine data providers to the UCL MRC CTU is also to be arranged in another application

Under agreement, **UCL** (UCL MRC CTU) would send a list of relevant patients securely to **PHE** (including linkers e.g. NHS number, DOB etc.)

PHE would extract the corresponding patient data and link STAMPEDE trial number alongside this (this would mean that the data sent on to us could be stripped of direct identifiers as we would no longer need identifiers to perform any linkage - other than via our STAMPEDE trial numbers).

This pseudonymised data (minus NHS number, DOB etc.) would be returned to the **UCL MRC CTU** to be used for analysis.

Frequency:

One – off	<input checked="" type="checkbox"/>
Periodic - monthly	<input type="checkbox"/>
Periodic - quarterly	<input type="checkbox"/>
Periodic - annually	<input type="checkbox"/>
Other	<input type="checkbox"/>

<p>Data already held for this project/purpose:</p> <p>Please include the dataset name, classification of the data (eg identifiable), the legal basis for processing, and the dataset period.</p>	<p><i>Clinical trial data (identifiable)</i></p> <p>STAMPEDE data: Held at the UCL MRC CTU under appropriate information governance with specific consent forms from all trial participants. This data has been linked to the corresponding initial pilot HES patient data.</p> <p><i>Data currently being applied for:</i></p> <ul style="list-style-type: none"> HES, RTDS, SACT, cancer registration: This data application must allow the linkage of the four PHE resources, and this data will directly be triangulated. This data application must also allow linkage of any of this data provided to us from PHE to ONS records that we are currently applying for from NHS Digital, and any other routine data source that we apply for in the future. The HES data from PHE will also be linked to new quarterly feeds of HES data from NHS Digital (once NHS Digital approval has been gained) to join to make a full dataset. Within the NHS Digital application, it will be explicitly stated that the HES data from NHS Digital will directly be linked to the HES data from PHE and triangulated to the RTDS, SACT, cancer registration, ONS etc. <p>We aim to link as many routine data sources as possible to build an accurate description of the STAMPEDE patients' case histories, to build clinical trial tools to enhance data collection, and therefore downstream enhance patients survival and quality of life.</p> <p>We need PHE to allow permission for us to access individual patients' STAMPEDE trial data and clinical note review data, when triggered from researching the PHE data</p> <p>Only processed data linked to trial ID will be permitted to leave the UCL MRC CTU</p> <ul style="list-style-type: none"> ONS data: Currently being applied for by UCL MRC CTU from NHS Digital
<p>Does this project involve patient contact (directly or indirectly through a clinical team/service provider)? If yes, please give details.</p>	<p>NO, there will be no direct contact of any patients on the basis of this data.</p>

Figure 46: Project summary and data specification excerpt from the PHE application that I completed for the STAMPEDE routine data.

8.1.4 PHE data variables



Annex A: Data Transfer Form

ODR reference:	ODR1718_094
Application title:	Using routine data to identify clinical trial outcomes within the STAMPEDE trial

Subject to the terms and conditions of the contract ODR1718_094, PHE agrees to transfer the Data as identified below:

Detailed Data Specification

<p>Cohort inclusion/exclusion criteria:</p> <p>Cohort inclusion criteria</p> <ul style="list-style-type: none"> Consented participants of the STAMPEDE trial - who have been diagnosed with C61 Consent remains extant <p>For matched records, the following case data will be disclosed to the data recipient in accordance with Tables 2-9:</p> <ul style="list-style-type: none"> Participants with a matched records in the NCRAS cancer registry All tumour records identified as C00-97 (inclusive of C44), D00 – 48 diagnosed between 2005-2016 For administrative datasets (HES inpatient, outpatient or A&E) the data will be indexed at 183 days (6 months) prior to the data to the participant's entry date on the trial) and be censored in accordance with the last available follow up date as per NCRAS data dictionary v3.6. The data will be censored in accordance to the information provided with each table.

For all fields below the subscript “_PS” indicates that the field has been pseudo-anonymised.

Table 1: For consented data subjects in the STAMPEDE trial cohort, the following information will be shared securely by the Data Recipient with PHE for cohort linkage

STAMPEDE TRIAL ID
NHS NUMBER
DATE OF BIRTH
DATE ENTERED TRIAL

Table 2: AV_PATIENT and AV_TUMOUR data will be provided to the Data Recipient in accordance with the cohort inclusion and exclusion criteria listed above.

STAMPEDEID
ETHNICITYNAME
TUMOURID_PS
DIAGNOSISDATEBEST
SITE_ICD10_O2
TRUSTCODE_FIRST_EVENT
DATE_FIRST_SURGERY
TRUSTCODE_FIRST_SURGERY

Table 3: Cancer Reg AV_TREATMENT data will be provided to the Data Recipient in accordance with the cohort inclusion and exclusion criteria .

STAMPEDEID
EVENTID_PS
TUMOURID_PS
DIAGNOSISDATEBEST
EVENTCODE
EVENTDATE
PROVIDERCODE
OPCSA_CODE
RADIOCODE
IMAGINGCODE
IMAGINGSITE
CHEMO_DRUG_ANON

Table 4: Linked HES IP data will be provided to the Data Recipient in accordance with the cohort inclusion and exclusion criteria: censored at 31/01/2018.

STAMPEDEID
EPIKEYANON_PS
ADMIDATE
DISDATE
EPISTART
EPIEND
EPIORDER
PROCEDURE
MAINSPEF
TRETSPEF
DIAG_mn
OPERTN_mn
OPDATE_mn

Table 5: Linked HES OP data will be provided to the Data Recipient in accordance with the cohort inclusion and exclusion criteria: censored at 31/01/2018.

STAMPEDEID

ATTENDEKEYANON_PS
APPTDATE
ATTENDED
MAINSPEF
TRETSPEF
PROCEDURE
DIAG_mn
OPERTN_mn

Table 6: Linked HES A&E data will be provided to the Data Recipient in accordance with the cohort inclusion and exclusion criteria: censored at 31/01/2018.

STAMPEDEID
AEKEYANON_PS
ARRIVALDATE
PROCEDURE
DIAG_mn
TREAT_mn
INVEST_mn

Table 7: SACT REGIMEN data will be provided to the Data Recipient in accordance with the cohort inclusion and exclusion criteria censored at 31/12/2017.

STAMPEDEID
MERGED_REGIMEN_ID_PS
MERGED_TUMOUR_ID_PS
PROGRAMME_NUMBER
REGIMEN_NUMBER
INTENT_OF_TREATMENT
ANALYSIS_GROUP
BENCHMARK_GROUP
HEIGHT_AT_START_OF_REGIMEN
WEIGHT_AT_START_OF_REGIMEN
PERF_STAT_START_OF_REG_CLEAN
COMORBIDITY_ADJUSTMENT
DATE_DECISION_TO_TREAT
START_DATE_OF_REGIMEN
CLINICAL_TRIAL
CHEMO_RADIATION
NUMBER_OF_CYCLES_PLANNED
PRIMARY_DIAGNOSIS
ORGANISATION_CODE_OF_PROVIDER
DATE_OF_FINAL_TREATMENT
REGIMEN_MOD_DOSE_REDUCTION
REGIMEN_MOD_TIME_DELAY
REGIMEN_MOD_STOPPED_EARLY

REGIMEN_OUTCOME_SUMMARY

Table 8 : SACT CYCLE data will be provided to the Data Recipient in accordance with the cohort inclusion and exclusion criteria censored at 31/12/2017.

STAMPEID
MERGED_REGIMEN_ID_PS
MERGED_CYCLE_ID_PS
CYCLE_NUMBER
START_DATE_OF_CYCLE
WEIGHT_AT_START_OF_CYCLE
PERF_STAT_START_OF_CYCLE_CLEAN
BENCHMARK_GROUP

Table 9 : SACT DRUG DETAIL data will be provided to the Data Recipient in accordance with the cohort inclusion and exclusion criteria censored at 31/12/2017.

STAMPEID
MERGED_REGIMEN_ID_PS
MERGED_CYCLE_ID_PS
MERGED_DRUG_DETAIL_ID_PS
DRUG_GROUP
ACTUAL_DOSE_PER_ADMINISTRATION
ADMINISTRATION_ROUTE
ADMINISTRATION_DATE

Table 10: RTDS EPISODE data will be provided to the Data Recipient in accordance with the cohort inclusion and exclusion criteria; censored at 31/03/2016.

STAMPEID
RADIOThERAPyEPISODEID_PS
ORGCODEPROVIDER
APPTDATE
ATTENDID_PS
DECISIONTOTREATDATE
TREATMENTSTARTDATE
RADIOThERAPyDIAGNOSISCD
RADIOThERAPyINTENT

Table 11: RTDS PRESCRIPTION data will be provided to the Data Recipient in accordance with the cohort inclusion and exclusion criteria; censored at 31/03/2016.

STAMPEID
RADIOThERAPyEPISODEID_PS
ORGCODEPROVIDER
APPTDATE

ATTENDID_PS
PRESCRIPTIONID_PS
RTTREATMENTREGION
RTTREATMENTANATOMICALSITE
RTPRESCRIBEDDOSE
PRESCRIBEDFRACTIONS
RTACTUALDOSE
ACTUALFRACTIONS
RTTREATMENTMODALITY

Table 12: RTDS EXPOSURE data will be provided to the Data Recipient in accordance with the cohort inclusion and exclusion criteria; censored at 31/03/2016.

STAMPEID
RADIOThERAPyEPISODEID_PS
ORGCODEPROVIDER
APPTDATE
ATTENDID_PS
PRESCRIPTIONID_PS
RADIOThERAPyFIELDID_PS
RADIOISOTOPE

Table 13: RTDS OPCLS data will be provided to the Data Recipient in accordance with the cohort inclusion and exclusion criteria; censored at 31/03/2016.

STAMPEID
ATTENDID_PS
APPTDATE
ORGCODEPROVIDER
PROCEDUREDATE

Figure 47: Copy of the document of the variables provided by PHE, upon application.

8.2 Chapter four

8.2.1 PRISMA guidelines (systematic literature review)

Section/topic	#	Checklist item	Reported?
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	✓
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	✓
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	✓ (chapter introduction)
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	✓
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	✗
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	✓
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	✓
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	✓
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	✓
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	✓
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	✓
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	✓
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	✓
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	✗
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	✓
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	✓

Section/topic	#	Checklist item	Reported?
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	✓
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	✓
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	✓
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	✓
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	✗
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	✓
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	✓
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	✓
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	✓
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	✓
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	✓ (section funding)

Figure 48: The PRISMA guidelines utilised in the systematic literature review

8.2.2 Content analyses

Criteria		Papers														
		Earle 2002	McClish 2003	Lamont 2006	Anaya 2011	Chubak 2012	Kimmick 2012	Nordstrom 2012	Liu 2013	Hagberg 2013	Lash 2014	Ricketts 2014	Ehrenstein 2015	Deshpande 2015	Haque 2015	Warren 2016
1	Participants	Patients not identified from an admin dataset (note review, registry)*	1	1	1	0	1	1	1	1	1	1	1	1	1	1
2		Participants - Event sample size for model development/validation >=50	0	1	0	1	1	1	1	1	0	1	1	0	1	1
3		Suitable patient exclusion criteria	0	0	1	1	0	0	0	0	0	0	1	0	1	0
4		Tabular patient characteristics	0	1	0	1	1	1	1	1	0	0	1	1	0	1
5	Participants	Patients in randomised controlled trial (RCT)	0	0	1	0	0	0	0	0	1	0	0	0	0	0
6		>one centre analysed over the whole study	0	1	NS	0	1	1	NS	1	1	0	0	0	1	1
7		Events compared to reference	1	1	1	1	1	1	1	1	1	1	1	1	1	0
8	Intervention / algorithm	Algorithm trained using sample (not a priori)	0	1	0	1	1	0	1	0	0	0	1	0	1	NS
9		Validation undertaken unseen dataset or cross-validation **	1	0	1	0	1	1	1	1	1	0	0	0	1	0
10		Algorithm detects outcome without pathology/lab reports (e.g. staging) or direct coding alone	1	1	0	1	1	1	1	NS	0	1	0	1	1	1
11	Outcome	Investigation into activity of variables (not necessarily in the final algorithm)	0	0	0	0	1	0	1	0	0	0	0	0	1	0
12		Methodology developed allowed timing of outcome to be identified	1	0	1	0	0	0	1	0	1	1	1	0	1	1
13		Algorithm feasible for use with regards to accuracy (>80% Se, Sp, PPV, NPV, where reported) (Indirect where present)	1	0	1	0	1	1	1	NS	1	0	0	0	0	NS
14	Quality	Reference data of high quality (e.g. note review, trial data, questionnaire, not registry data)	1	0	1	1	1	1	1	1	1	1	1	1	1	0
15		Table/figure comparing results algorithm vs. reference data	1	1	1	1	1	0	1	0	1	1	1	1	1	0
16		Indexed by PubMed	1	1	1	1	1	1	1	1	1	1	1	1	1	1
17		Algorithm variables published (or summary)	1	0	1	1	1	1	0	0	1	1	1	0	1	1
TOTAL SCORE: /17		10	9	11	10	14	11	13	9	7	12	10	11	7	14	8
% score		58.8	52.9	64.7	58.8	82.4	64.7	76.5	52.9	41.2	70.6	58.8	64.7	41.2	82.4	47.1

Table 79: Part 1) Content analyses, assessing weakness and bias. * Registry = considered accurate to identify patients but not to identify outcomes; ** If the algorithm was developed a priori (not trained using a sample) then the study was considered to validate on an unseen dataset.

	Criteria	Papers												SUMMARY	
		Livaudais 2016	Nordstrom 2016	Joshy 2016	Hassett 2017	Gupta 2017	Ritzwoller 2017	Wong 2018	Rasmussen 2018	Uno 2018	Xu 2019	Xu 2019	Rasmussen 2019	Sum	% of papers
1	Patients not identified from an admin dataset (note review, registry)*	0	1	1	1	1	1	1	1	1	NS	1	1	24	89
2	Participants - Event sample size for model development/Validation >=50	0	1	1	1	1	1	0	1	1	1	1	1	21	78
3	Suitable patient exclusion criteria	0	0	0	0	1	0	NS	0	NS	0	0	0	5	19
4	Tabular patient characteristics	1	1	1	1	1	0	0	1	0	1	1	1	19	70
5	Patients in randomised controlled trial (RCT)	0	0	0	0	0	0	0	0	0	0	0	0	2	7
6	>one centre analysed over the whole study	0	1	1	1	1	1	0	1	1	0	1	1	17	63
7	Events compared to reference	1	1	1	1	1	1	1	1	1	1	1	1	26	96
8	Algorithm trained using sample (not a priori)	NS	1	1	1	1	1	1	1	1	1	1	1	18	67
9	Validation undertaken using unseen dataset or cross-validation**	0	1	1	1	0	1	0	0	1	1	1	0	15	56
10	Algorithm detects outcome without pathology/lab reports (e.g. staging) or direct coding alone	1	1	1	1	1	1	1	0	1	1	1	0	21	78
11	Investigation into activity of variables (not necessarily in the final algorithm)	0	0	0	1	0	1	0	0	1	1	1	0	8	30
12	Methodology developed allowed timing of outcome to be identified	0	0	0	1	1	1	1	1	1	0	0	1	15	56
13	Outcome	1	0	0	0	0	0	0	0	NS	0	1	1	9	33
14	Reference data of high quality (e.g. note review, trial data, questionnaire, not registry data)	1	1	1	1	1	1	1	0	1	1	1	0	23	85
15	Quality	0	1	0	0	0	1	1	1	1	0	0	1	17	63
16	Table/figure comparing results algorithm vs. reference data	1	1	1	1	1	1	1	1	1	1	1	1	27	100
17	Indexed by PubMed	1	0	1	1	1	1	1	0	0	1	1	0	19	70
TOTAL SCORE: /17		7	11	11	13	12	13	9	9	12	10	13	10		
% score		41.2	64.7	64.7	76.5	70.6	76.5	52.9	52.9	70.6	58.8	76.5	58.8		

Table 80: Part 2) Content analyses, assessing weakness and bias. * Registry = considered accurate to identify patients but not to identify outcomes; ** If the algorithm was developed a priori (not trained using a sample) then the study was considered to validate on an unseen dataset.

	Criteria	Do the studies presented in this thesis chapter meet the weakness and bias criteria?
Participants	Patients not identified from an admin dataset (note review, registry)	Y
	Participants - Event sample size for model development/validation >=50	Y
	Suitable patient exclusion criteria	Y
	Tabular patient characteristics	Y
	Patients in randomised controlled trial (RCT)	Y
	>one centre analysed over the whole study	Y
Intervention / algorithm	Events compared to reference	Y
	Algorithm trained using sample (not a priori)	Y
	Validation undertaken unseen dataset or cross-validation	Y
	Algorithm detects outcome without pathology/lab reports (e.g. staging) or direct coding alone	Y
	Investigation into activity of variables (not necessarily in the final algorithm)	Y
Outcome	Methodology developed allowed timing of outcome to be identified	Y
	Algorithm feasible for use with regards to accuracy (>80% Se, Sp, PPV, NPV, where reported) (indirect where present)	NA (novel endpoint developed)
	Reference data of high quality (e.g. note review, trial data, questionnaire, not registry data)	Y
Quality	Table/figure comparing results algorithm vs. reference data	Y
	Indexed by PubMed	Y (when ICTMC abstract published)
	Algorithm variables published (or summary)	Y

Table 81: Assessment of this thesis study in comparison to the developed weakness and bias criteria from the systematic review.
NA = not appropriate.

8.2.3 Interval analyses

ID		Weeks (number of peaks)																							
		1			3			4				6				8				10				12	
	T+	F+	F-	T+	F+	F-	No. peaks	T+	F+	F-	No. peaks	T+	F+	F-	No. peaks	T+	F+	F-	No. peaks	T+	F+	F-	T+	F+	F-
1	Not assessed due to 4-weekly clustering noise			Not assessed due to 4-weekly clustering noise			10	3	7	0	8	3	5	0	4	3	1	0	4	3	1	0	Too broad to be clinically relevant		
2							Not investigated further				3	2	1	0	2	2	0	0	1	1	0	1			
3							2	1	1	0	1	1	0	0	1	1	0	0	1	1	0	0			
4							Not investigated further				3	2	1	0	3	2	1	0	2	2	0	0			
5											12	3	9	0	8	3	5	0	6	3	3	0			
Totals (number of peaks, T+, F+, F-)							27	11	16	0	18	11	7	0	14	10	4	1							

Table 82: Interval analysis conducted to find the optimum time interval to identify outcomes from the prostate cancer-related event clustering.
T+ = true positives; F+ = false positives; F- = false negatives.

Weekly interval	Reason for inclusion/exclusion
1	See below*
3	See below*
4	See below*
6	Too many false positives were present, but all events (11 events) were identified
8	All events were identified (11 events) but there were less false positives than with the 6-weekly clustering
10	One less true positive was identified and these broader intervals are less clinically relevant
12	A 12 week interval was determined to be too broad to be clinically relevant

Table 83: The rationale behind the 8-weekly event clustering that was chosen for the algorithm.

* = Upon analysing two patients by clustering events into 4-weekly intervals, it was clear that intervals ≤ 4 would not be feasible, too many false positives were present (hence ≤ 1 and ≤ 3 was not considered).

8.2.4 Threshold analyses

Patient	Total events	Threshold = 6 events				Threshold = 5 events				
		Algorithm detected	False negatives	% detected (sensitivity)	False positives	PPV	Algorithm detected	False negatives	% detected (Se)	False positives
1	3	2	1	66.7	0	75.0	3	0	100	0
2	3	3	0	100	3		3	0	100	4
3	1	1	0	100	0		1	0	100	0
4	2	2	0	100	0		2	0	100	0
5	4	4	0	100	1		4	0	100	1
Total	13	12	1	92.3	4		13	0	100.0	5

Table 84: Further detail into the number of events detected and missed at the five and six event thresholds.

8.2.5 Outcome analyses

Event	Detection	Progression outcome (weeks)						
		Routine data model output				Note review		Trial
		Event interval	Interval midpoint	Event interval	Interval midpoint	Event date	Interval midpoint	
		Threshold 6 episodes		Threshold 5 episodes				
1	Bone scan	Missed	Missed	[8,16)	12	16	20	Y
2	Bone scan	[64-72)	68	[64-72)	68	66	68	N
3	Bone scan	[96-104)	100	[96-104)	100	102	100	Y
4	Bone scan	[72-80)	76	[72-80)	76	80	84	Y
5	SRE-symptomatic bone pain	[136-144)	140	[136-144)	140	138	140	Y
6	SRE - radiotherapy	[216-224]	220	[216-224]	220	218	220	Y
7	MRI scan	[8-16)	12	[8-16)	12	11	12	Y
8	Bone scan	[16-24)	20	[16-24)	20	17	20	Y
9	CT scan	[32-40)	36	[32-40)	36	36	36	Y
10	Bone scan	[96-104)	100	[96-104)	100	92	92	Y
11	MRI	[128-136)	132	[128-136)	132	118	116	N
12	Bone scan	[152-160)	156	[152-160)	156	154	156	N
13	MRI	[176-184)	180	[176-184)	180	166	164	N

Table 85: Comparative analyses for the five and six event thresholds for the routine data model, compared to the note review (also grouped into the corresponding 8-week interval) and the STAMPEDE trial.

Event	Difference routine data midpoint & note review (w \pm)	Number intervals difference (\pm)	Difference routine data midpoint & note review (w \pm)	Number intervals difference (\pm)
	Threshold = 6 events		Threshold = 5 events	
1	NA	NA	-4	-1
2	2	0	2	0
3	-2	0	-2	0
4	-4	-1	-4	-1
5	2	0	2	0
6	2	0	2	0
7	1	0	1	0
8	3	0	3	0
9	0	0	0	0
10	8	1	8	1
11	14	2	14	2
12	2	0	2	0
13	14	2	14	2

Table 86: The difference in the model detected-events (shown in table 85) and the note review detected events, using the 5 and 6 event thresholds and the 8-week intervals.

8.2.6 SRE analysis

ID	SRE (w = week)	PDE cluster?	Data found summary	Note review	HES	HES + inferred	Trial data	Total
5	RT w118	No	HES, NR, CRF	1	1	1	1	6 = RT (N=5), SCC (N=1)
	RT w154	Yes	HES (found 800 HES), NR	1	0	1	0	
	RT w166	No	HES, NR, CRF	1	1	1	1	
	SCC w166	No	NR, CRF, HES	1	1	1	1	
	RT w188	Yes	HES, NR	1	1	1	0	
2	RT w189	Yes	NR, HES (found 800 HES)	1	0	1	0	1 = RT (N=1)
	RT w219	No	HES, NR (found 800 HES)	1	0	1	0	
3	SCC w13	Yes	HES, NR, CRF	1	1	1	1	2 = SCC (N=1), surgery (N=1)
	Surgery w13	Yes	HES (as surgery), NR, CRF	1	0	1	1	
	RT w20	No	NR, CRF, HES (found 800 HES)	1	0	1	1	
1	RT w56	Yes	NR, CRF, HES (found 800 HES)	1	0	1	1	5 = RT (N=3), pathological fracture (N=1), SCC (N=1)
	RT w78	No	NR, CRF*	1	0	0	1	
	Pathological fracture (w55 XR)	No	NR	1	0	0	0	
	SCC (w116 wheelchair)	Yes	NR	1	0	0	0	6 = RT (N=5), SCC (N=1)
	RT w21	Yes	NR, CRF	1	0	0	1	
	RT w24	No	NR, HES (found 800 HES)	1	0	1	0	
	RT SCC w35	Yes	NR, HES	1	1	1	0	
4	SCC w35	Yes	NR, HES, CRF	1	1	1	1	6 = RT (N=5), SCC (N=1)
	RT pelvis w38	Yes	NR, HES	1	1	1	0	
	RT w42	No	NR, HES (found 800 HES)	1	0	1	0	
6	None	None	N/A	0	0	0	0	(N=0)
				20	8	16	10	

Table 87: Table calculating the accuracy of the routine data to identify SREs. RT = radiotherapy; SCC = spinal cord compression; Blue = note review, HES and trial data; yellow = note review and HES; green = note review and trial data; pink = note review only.

8.2.7 Outcomes

Data/Outcome	Non-chemotherapy arms, ABDF (n/N) (%)	Chemotherapy arms, CE (n/N) (%)
Note review PFS	24/29 (83%)	9/15 (60%)
STAMPEDE PFS	19/29 (66%)	5/15 (33%)
STAMPEDE MFS	19/29 (66%)	6/15 (40%)
STAMPEDE FFS	24/29 (83%)	9/15 (60%)
HES AFS	21/29 (72%)	6/15 (40%)
SREs		
STAMPEDE SRE	13/29 (45%)	5/15 (33%)
HES SRE	5/29 (17%)	2/15 (13%)

Table 88: The number of cohort two outcomes detected, split by treatment, for each endpoint.
n = total number of events; *N* = total number of patients.

Data/Outcome	Non-chemotherapy arms, ABDF (n/N) (%)	Chemotherapy arms, CE (n/N) (%)
STAMPEDE PFS	46/62 (74%)	10/31 (32%)
STAMPEDE MFS	44/62 (71%)	12/31 (39%)
STAMPEDE FFS	50/62 (81%)	17/31 (55%)
HES AFS	50/62 (81%)	16/31 (52%)
SREs		
STAMPEDE SRE	27/62 (44%)	7/31 (23%)
HES SRE	11/62 (18%)	4/31 (13%)

Table 89: The number of cohort three outcomes detected, split by treatment, for each endpoint.
n = total number of events; *N* = total number of patients.

Data/Outcome	Non-chemotherapy standard-of-care arms, A (n/N) (%)	Chemotherapy arms, CE (n/N) (%)
STAMPEDE PFS	417/855 (48.8%)	337/840 (40.1%)
STAMPEDE MFS	400/855 (46.8%)	348/840 (41.4%)
STAMPEDE FFS	567/855 (66.3%)	461/840 (54.9%)
HES AFS	439/855 (51.3%)	394/840 (46.9%)
SREs		
STAMPEDE SRE	249/855 (29.1%)	162/840 (19.3%)
HES SRE	189/855 (22.1%)	134/840 (16.0%)

Table 90: The number of cohort four outcomes detected, split by treatment, for each endpoint.
n = total number of events; *N* = total number of patients.

8.2.8 SRE scatter plot

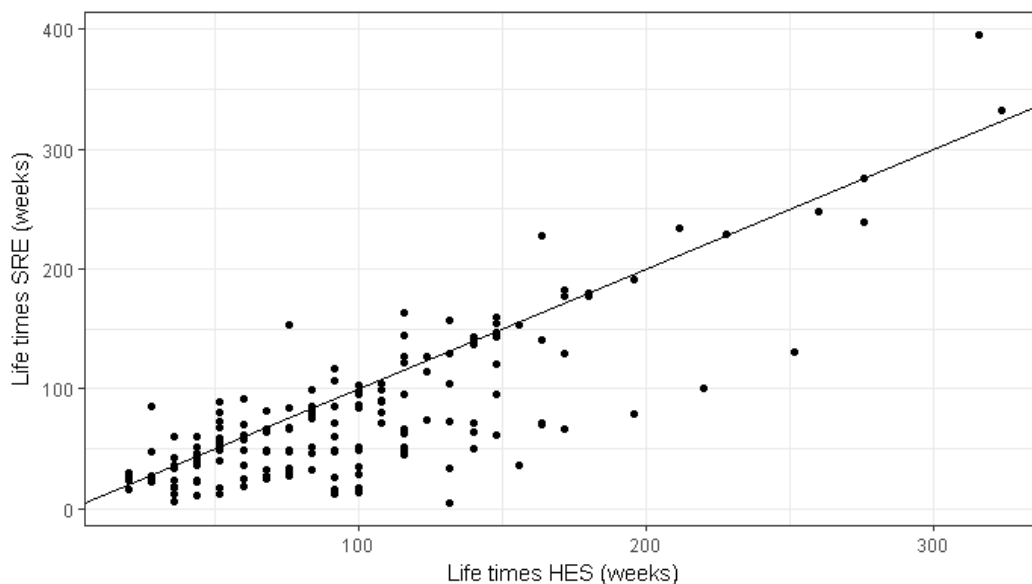


Figure 49: A scatter plot of the correlation between the HES detected SREs and the STAMPEDE detected SREs.

Includes only patients where a HES outcome and a STAMPEDE outcome were both identified. The black line through the origin is a line to illustrate perfect correlation.

8.3 Chapter five

8.3.1 Time to correct treatment (TTCT) event coding

Flexible cystoscopy	
M459	Unspecified diagnostic endoscopic examination of bladder
Rigid cystoscopy (TURBT)	
M421	Endoscopic resection of lesion of bladder
Surgery to bladder	
M341	Cystoprostatectomy
M342	Cystourethrectomy
M343	Cystectomy NEC
M344	Simple cystectomy
M348	Other specified total excision of bladder
M349	Unspecified total excision of bladder
M358	Other specified partial excision of bladder
M359	Unspecified partial excision of bladder
X141	Total exenteration of pelvis
X142	Anterior exenteration of pelvis
X148	Other specified clearance of pelvis
X149	Unspecified clearance of pelvis
Radiotherapy: Radiotherapy planning	
X671	Preparation for intensity modulated radiation therapy
X672	Preparation for total body irradiation
X673	Preparation for hemi body irradiation
X674	Preparation for simple radiotherapy with imaging and dosimetry
X675	Preparation for simple radiotherapy with imaging and simple calculation
X676	Preparation for superficial radiotherapy with simple calculation

X677	Preparation for complex conformal radiotherapy
X678	Other specified preparation for external beam radiotherapy
X679	Unspecified preparation for external beam radiotherapy
Y921	Technical support for preparation for radiotherapy
Y928	Other specified support for preparation for radiotherapy
Y929	Unspecified support for preparation for radiotherapy
Radiotherapy delivery	
Y902	Radiotherapy NEC
X651	Delivery of a fraction of total body irradiation
X652	Delivery of a fraction of intracavitary radiotherapy
X653	Delivery of a fraction of interstitial radiotherapy
X654	Delivery of a fraction of external beam radiotherapy NEC
X658	Other specified radiotherapy delivery
X659	Unspecified radiotherapy delivery
Radiotherapy type	
Y911	Megavoltage treatment for complex radiotherapy
Y912	Megavoltage treatment for simple radiotherapy
Y913	Superficial or orthovoltage treatment for radiotherapy
Y914	Megavoltage treatment for adaptive radiotherapy
Y915	Megavoltage treatment for hypofractionated stereotactic radiotherapy
Y918	Other specified external beam radiotherapy
Y919	Unspecified external beam radiotherapy
Chemotherapy: Chemotherapy procurement	
X701	Procurement of drugs for chemotherapy for neoplasm for regimens in Band 1
X702	Procurement of drugs for chemotherapy for neoplasm for regimens in Band 2
X703	Procurement of drugs for chemotherapy for neoplasm for regimens in Band 3
X704	Procurement of drugs for chemotherapy for neoplasm for regimens in Band 4
X705	Procurement of drugs for chemotherapy for neoplasm for regimens in Band 5
X708	Other specified procurement of drugs for chemotherapy for neoplasm in Bands 1-5
X709	Unspecified procurement of drugs for chemotherapy for neoplasm in Bands 1-5
X711	Procurement of drugs for chemotherapy for neoplasm for regimens in Band 6
X712	Procurement of drugs for chemotherapy for neoplasm for regimens in Band 7
X713	Procurement of drugs for chemotherapy for neoplasm for regimens in Band 8
X714	Procurement of drugs for chemotherapy for neoplasm for regimens in Band 9
X715	Procurement of drugs for chemotherapy for neoplasm for regimens in Band 10
X718	Other specified procurement of drugs for chemotherapy for neoplasm in Bands 6-10
X719	Unspecified procurement of drugs for chemotherapy for neoplasm in Bands 6-10
Chemotherapy delivery	
X352	Intravenous chemotherapy
X373	Intramuscular chemotherapy
X384	Subcutaneous chemotherapy
X721	Delivery of complex chemotherapy for neoplasm including prolonged infusional treatment at first attendance
X722	Delivery of complex parenteral chemotherapy for neoplasm at first attendance
X723	Delivery of simple parenteral chemotherapy for neoplasm at first attendance
X724	Delivery of subsequent element of cycle of chemotherapy for neoplasm
X728	Other specified delivery of chemotherapy for neoplasm
X729	Unspecified delivery of chemotherapy for neoplasm
X731	Delivery of exclusively oral chemotherapy for neoplasm
X738	Other specified delivery of oral chemotherapy for neoplasm
X739	Unspecified delivery of oral chemotherapy for neoplasm
X748	Other specified other chemotherapy drugs
X749	Unspecified other chemotherapy drugs
Y123	Electrochemotherapy to lesion of organ NOC

Table 91: The codes used to identify the events to calculate time to correct treatment (TTCT) from the routine data.

8.3.2 Data quality event coding

Code	Value
<u>Cystoscopy (all)</u>	
M451	Diagnostic endoscopic examination of bladder and biopsy of lesion of bladder NEC
M452	Diagnostic endoscopic examination of bladder and biopsy of lesion of prostate NEC
M453	Diagnostic endoscopic examination of bladder and biopsy of lesion of bladder using rigid cystoscope
M454	Diagnostic endoscopic examination of bladder and biopsy of lesion of prostate using rigid cystoscope
M455	Diagnostic endoscopic examination of bladder using rigid cystoscope
M458	Other specified diagnostic endoscopic examination of bladder
M459	Unspecified diagnostic endoscopic examination of bladder
M421	Endoscopic resection of lesion of bladder
M422	Endoscopic cauterisation of lesion of bladder
M423	Endoscopic destruction of lesion of bladder NEC
M428	Other specified endoscopic extirpation of lesion of bladder
M429	Unspecified endoscopic extirpation of lesion of bladder
M411*	Open extirpation of lesion of bladder
M448	Other specified other therapeutic endoscopic operations on bladder
<u>BCG</u>	
M494	Introduction of therapeutic substance into bladder
M495	Injection of therapeutic substance into bladder wall
E952	Administration of Bacillus Calmette-Guerin vaccine
X448	Other specified administration of vaccine
X449	Unspecified administration of vaccine
M479	Unspecified urethral catheterisation of bladder (only in combination with the below)
M479 + X721	Delivery of complex chemotherapy for neoplasm including prolonged infusional treatment at first attendance
M479 + X722	Delivery of complex parenteral chemotherapy for neoplasm at first attendance
M479 + X723	Delivery of simple parenteral chemotherapy for neoplasm at first attendance
M479 + X724	Delivery of subsequent element of cycle of chemotherapy for neoplasm
M479 + X728	Other specified delivery of chemotherapy for neoplasm
M479 + X729	Unspecified delivery of chemotherapy for neoplasm
M479 + X748	Other specified other chemotherapy drugs
M479 + X749	Unspecified other chemotherapy drugs
<u>Surgery to bladder</u>	
M341	Cystoprostatectomy
M342	Cystourethrectomy
M343	Cystectomy NEC
M344	Simple cystectomy
M348	Other specified total excision of bladder
M349	Unspecified total excision of bladder
M358	Other specified partial excision of bladder
M359	Unspecified partial excision of bladder
M418	Other specified other open operations on bladder
M419	Unspecified other open operations on bladder
M488	Other specified operations on bladder
M489	Unspecified operations on bladder
M498	Other specified other operations on bladder
M499	Unspecified other operations on bladder
X141	Total exenteration of pelvis

X142	Anterior exenteration of pelvis
X148	Other specified clearance of pelvis
X149	Unspecified clearance of pelvis
Radiotherapy	
Radiotherapy planning	
X671	Preparation for intensity modulated radiation therapy
X672	Preparation for total body irradiation
X673	Preparation for hemi body irradiation
X674	Preparation for simple radiotherapy with imaging and dosimetry
X675	Preparation for simple radiotherapy with imaging and simple calculation
X676	Preparation for superficial radiotherapy with simple calculation
X677	Preparation for complex conformal radiotherapy
X678	Other specified preparation for external beam radiotherapy
X679	Unspecified preparation for external beam radiotherapy
Y921	Technical support for preparation for radiotherapy
Y928	Other specified support for preparation for radiotherapy
Y929	Unspecified support for preparation for radiotherapy
Radiotherapy delivery	
Y902	Radiotherapy NEC
X651*	Delivery of a fraction of total body irradiation
X652*	Delivery of a fraction of intracavitary radiotherapy
X653*	Delivery of a fraction of interstitial radiotherapy
X654	Delivery of a fraction of external beam radiotherapy NEC
X658	Other specified radiotherapy delivery
X659	Unspecified radiotherapy delivery
Radiotherapy type	
Y911	Megavoltage treatment for complex radiotherapy
Y912	Megavoltage treatment for simple radiotherapy
Y913	Superficial or orthovoltage treatment for radiotherapy
Y914	Megavoltage treatment for adaptive radiotherapy
Y915	Megavoltage treatment for hypofractionated stereotactic radiotherapy
Y918	Other specified external beam radiotherapy
Y919	Unspecified external beam radiotherapy
Chemotherapy	
Chemotherapy procurement	
X701	Procurement of drugs for chemotherapy for neoplasm for regimens in Band 1
X702	Procurement of drugs for chemotherapy for neoplasm for regimens in Band 2
X703	Procurement of drugs for chemotherapy for neoplasm for regimens in Band 3
X704	Procurement of drugs for chemotherapy for neoplasm for regimens in Band 4
X705	Procurement of drugs for chemotherapy for neoplasm for regimens in Band 5
X708	Other specified procurement of drugs for chemotherapy for neoplasm in Bands 1-5
X709	Unspecified procurement of drugs for chemotherapy for neoplasm in Bands 1-5
X711	Procurement of drugs for chemotherapy for neoplasm for regimens in Band 6
X712	Procurement of drugs for chemotherapy for neoplasm for regimens in Band 7
X713	Procurement of drugs for chemotherapy for neoplasm for regimens in Band 8
X714	Procurement of drugs for chemotherapy for neoplasm for regimens in Band 9
X715	Procurement of drugs for chemotherapy for neoplasm for regimens in Band 10
X718	Other specified procurement of drugs for chemotherapy for neoplasm in Bands 6-10
X719	Unspecified procurement of drugs for chemotherapy for neoplasm in Bands 6-10
Chemotherapy delivery	
X352	Intravenous chemotherapy
X373*	Intramuscular chemotherapy
X384*	Subcutaneous chemotherapy
X721	Delivery of complex chemotherapy for neoplasm including prolonged infusional treatment at first attendance
X722	Delivery of complex parenteral chemotherapy for neoplasm at first attendance
X723	Delivery of simple parenteral chemotherapy for neoplasm at first attendance
X724	Delivery of subsequent element of cycle of chemotherapy for neoplasm

X728	Other specified delivery of chemotherapy for neoplasm
X729	Unspecified delivery of chemotherapy for neoplasm
X731*	Delivery of exclusively oral chemotherapy for neoplasm
X738*	Other specified delivery of oral chemotherapy for neoplasm
X739*	Unspecified delivery of oral chemotherapy for neoplasm
X748	Other specified other chemotherapy drugs
X749	Unspecified other chemotherapy drugs
Y123*	Electrochemotherapy to lesion of organ NOC
Chemotherapy (other)	
X358	Other specified other intravenous injection
X292	Continuous intravenous infusion of therapeutic substance NEC
X293*	Continuous subcutaneous infusion of therapeutic substance NEC
X298	Other specified continuous infusion of therapeutic substance
X299	Unspecified continuous infusion of therapeutic substance
X281	Intermittent intravenous infusion of therapeutic substance
X282*	Intermittent subcutaneous infusion of therapeutic substance
X288	Other specified intermittent infusion of therapeutic substance
X289	Unspecified intermittent infusion of therapeutic substance
X308	Other specified injection of therapeutic substance
X309	Unspecified injection of therapeutic substance
X391*	Oral administration of therapeutic substance
X398*	Other specified other route of administration of therapeutic substance
X399*	Unspecified other route of administration of therapeutic substance
X353*	Intravenous immunotherapy
X374*	Intramuscular immunotherapy
X385*	Subcutaneous immunotherapy
Last follow-up censor event (attended inpatient or outpatient event)	
101	Urology
370	Medical oncology
800	Clinical oncology

Table 92: Codes identified by the algorithm to detect events for the data quality validation.
(* included to identify miscoded procedures).

A: Word count tables, figures, captions: 16,927

B: Appendix word count: 4,595

C: Reference word count: 6,224

D: Total word count: 92,931

Final word count (exclusive of appendices, figures/tables, captions and references but inclusive of all pre-text, in-text reference numbering and caption cross-references)

$$D - (A + B + C) = 92,931 - (16,927 + 4,595 + 6,224 = 27,746): \underline{65,185}$$